EARTHQUAKE PREDICTION SYSTEM

A main Project Report submitted in the partial fulfillment of

the requirements for the award of the degree

BACHELOR OFTECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

B. Bala Murali Krishna	(19471A0571)
M. Aditya	(19471A05A3)
R. Vijay Kumar	(19471A03B4)

Under the esteemed guidance of

Sd. Rizwana, M. Tech (Ph. D)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

NARASARAOPETA ENGINEERING COLLEGE:

NARASARAOPET (AUTONOMOUS)

(Affiliated to J.N.T.U.Kakinada ,Approved by AICTE&Acredited by NBA)

Accredited by NAAC with A+ Grade and NBA under
Cycle-1 NIRF rank in the band of 251-320 and an ISO
9001-2015 Certified
Approved by AICTE, New Delhi, Permanently Affiliated to JNTUK,
KakinadaKOTAPPAKONDA ROAD, YALAMANDA VILLAGE, NARASARAOPET-522601

2022-2023

NARASARAOPETA ENGINEERING COLLEGE: NARASARAOPETA (AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the main project entitled " **EARTHQUAKE PREDICTION SYSTEM"** is a Bonafide work done by B. Bala Murali Krishna(19471A0571), M. Aditya(19471A05A3), R.Vijay Kumar(19471A05B4)" in partial fulfilment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in the department of **COMPUTER SCIENCE AND ENGINEERING during 2022-2023.**

PROJECT GUIDE

Sd. Rizwana, M. Tech (Ph. D)

PROJECT CO-ORDINATOR

Dr.M. Sireesha, M. Tech, Ph. D

HEAD OF THE DEPARTMENT

Dr.S.N. Tirumala Rao, M. Tech, Ph. D

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We wish to express our thanks to carious personalities who are responsible for the completion of the project. We are extremely thankful to our beloved chairman sir **M.V. Koteswara Rao**, **B.sc** who took keen interest in us in every effort throughout this course. We owe out gratitude to our principal **Dr. M.Sreenivasa Kumar, M.Tech,Ph.D (UK),MISTE,FIE(1)** for his kind attention and valuable guidance throughout the course.

We express our deep felt gratitude to **Dr.S.N.Tirumala Rao, M .Tech, Ph. D** H.O.D. CSEdepartment and our guide **Sd. Rizawana, M.Tech (Ph. D)** of CSE department whose valuableguidance and unstinting encouragement enable us to accomplish our project successfully in time.

We extend our sincere thanks to Dr.**M. Sireesha, M.Tech, Ph.D** Associate Professor coordinator of the project for extending her encouragement. Their profound knowledge and willingness have been a constant source of inspiration for us throughout this project work.

We extend our sincere thanks to all other teaching and non-teaching staff to department for their cooperation and encouragement during our B.Tech degree. We have no words to acknowledge the warm affection, constant inspiration and encouragement that we receive from our parents. We affectionately acknowledge the encouragement received from our friends those who involved in giving valuable suggestions had clarifying out all doubts which had really helped us in successfully completing our project.

By

B. Bala Murali Krishna (19471A0571)

M. Aditya (19471A05A3)

R. Vijay Kumar (19471A05B4)

ABSTRACT

During this study, earthquake prediction was performed, by training different Machine Learning models on seismic and acoustic data collected from a laboratory micro -earthquake simulation. Prediction has been made by extracting 40 statistical features, such as no. of peaks, time to failure etc. from the 'single-feature' acoustic data, which was basically in the form of a time series. During this research, six machine learning techniques including Linear Regression, Support Vector Machine, Random Forest Regression, Case Based Reasoning, XGBoost and Light Gradient Boosting Mechanism are separately applied and accuracies in the training and testing datasets were compared to pick out the best model. Furthermore, the evaluation of accuracy is another step taken into account for analysing the result. The above methods for predicting earthquake magnitude yield significant and encouraging results,.

NEC NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

INSTITUTE VISION AND MISSION

INSTITUTION VISION

To emerge as a Centre of excellence in technical education with a blend of effective student centric teaching learning practices as well as research for the transformation of lives and community,

INSTITUTION MISSION

M1: Provide the best class infra-structure to explore the field of engineering and research

M2: Build a passionate and a determined team of faculty with student centric teaching, imbibing experiential, innovative skills

M3: Imbibe lifelong learning skills, entrepreneurial skills and ethical values in students for addressing societal problems.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VISION OF THE DEPARTMENT

To become a centre of excellence in nurturing the quality Computer Science & Engineering professionals embedded with software knowledge, aptitude for research and ethical values to cater to the needs of industry and society.

MISSION OF THE DEPARTMENT

The department of Computer Science and Engineering is committed to

M1: Mould the students to become Software Professionals, Researchers and Entrepreneurs byproviding advanced laboratories.

M2: Impart high quality professional training to get expertize in modern software tools and technologies to cater to the real time requirements of the Industry.

M3: Inculcate team work and lifelong learning among students with a sense of societal andethical responsibilities.

NARASARAOPETA ENGINEERING COLLEGE (AUTONOMOUS)

Program Specific Outcomes (PSO's)

PSO1: Apply mathematical and scientific skills in numerous areas of Computer Scienceand Engineering to design and develop software-based systems.

PSO2: Acquaint module knowledge on emerging trends of the modern era in Computer Science and Engineering

PSO3: Promote novel applications that meet the needs of entrepreneur, environmental and social issues.

NEC NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Educational Objectives (PEO’s)

The graduates of the programme are able to:

PEO1: Apply the knowledge of Mathematics, Science and Engineering fundamentals toidentify and solve Computer Science and Engineering problems.

PEO2: Use various software tools and technologies to solve problems related to academia,industry and society.

PEO3: Work with ethical and moral values in the multi-disciplinary teams and can communicate effectively among team members with continuous learning.

PEO4: Pursue higher studies and develop their career in software industry.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Outcomes

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineeringproblems.

2. Problem analysis: Identify, formulate, research literature, and analyzecomplex engineering problems reaching substantiated conclusions using firstprinciples of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complexengineering problems and design system components or processes that meetthe specified needs with appropriate consideration for the public health andsafety, and the cultural, societal, and environmental considerations.

4.Conduct investigations of complex problems: Use research-basedknowledge and research methods including design of experiments, analysis andinterpretation of data, and synthesis of the information to provide validconclusions.

5.Modern tool usage: Create, select, and apply appropriate techniques,resources, and modern engineering and IT tools including prediction andmodeling to complex engineering activities with an understanding of thelimitations.

6.The engineer and society: Apply reasoning informed by the contextualknowledge to assess societal, health, safety, legal and cultural issues and theconsequent responsibilities relevant to the professional engineering practice.

Environment and sustainability: Understand the impact of theprofessional engineering solutions in societal and environmental contexts,and demonstrate the knowledge of, and need for sustainable development.

Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

Individual and team work: Function effectively as an individual, and as a member or leaderin diverse teams, and in multidisciplinary settings.

Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receiveclear instructions.

Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Project Course Outcomes (CO'S):

CO425.1: Analyse the System of Examinations and identify the problem.

CO425.2: Identify and classify the

requirements.**CO425.3:** Review the

Related Literature **CO425.4:**

Design and Modularize the project

CO425.5: Construct, Integrate, Test and Implement the Project.

CO425. 6: Prepare the project Documentation and present the Report using appropriatemethod.

Course Outcomes – Program Outcomes mapping

	P O 1	P O 2	P O 3	PO 4	P O 5	P O 6	P O 7	P O 8	P O 9	PO 10	PO 11	PO 12	PS 01	PSO 2	PSO 3
C42 5.1		√											√		
C42 5.2	√		√		√								√		
C42 5.3				√		√	√	√					√		
C42 5.4			√			√	√	√					√	√	
C42 5.5					√	√	√	√	√	√	√	√	√	√	√
C42 5.6									√	√	√		√	√	

INDEX

1.Introduction

1.1 Introduction

Earthquake is one of the devastating events in natural hazards that causes great casualties and property damage every day in the world since that it is hard to predict. With the increasing amount of earthquake datasets collected, many researchers try to solve the task of predicting the earthquake in the future time. Earthquake prediction is to estimate the time, location and magnitude of the future earthquake, which is one of the theoretical foundation of geophysics, geology, computer science and so on. With the development of data mining techniques, a large number of scholars have devoted to discover the earthquake patterns from seismic time series based on various feature extraction methods and achieved some success.

The prediction of earthquake as a whole has proved to be a challenge which is essentially impossible. With modern computing power, machine learning techniques, and a significantly narrowed emphasis, however, this important role may possibly be made some headway. To this end, predict the occurrence of earthquakes using seismic signal.

Earthquake prediction research projects in a variety of countries are reviewed in accordance with achievements in various disciplines involved in earthquake prediction science, i.e. geodesic work, observation of tide gages, continuous observation of crustal m Acoustic / seismic precursors to failure seem to be an almost common material phenomenon. For example, failure in granular materials is often followed by impulsive acoustic / seismic precursors, many of which are very small. Precursors are found in laboratory faults and corresponding earthquakes are commonly but not routinely detected. We claim that the magnitudes of seismic precursors can be very small, and so often go unrecorded or unidentified.

An earthquake consists of many individual elastic waves. The functioning of EWS is based on the principle that these waves have travelled from the epicenter to recombine at the recording site as a function of their respective velocities, focal distances, and propagation paths. Body waves propagate within a body of rock and appear in the first arrival. When a natural disaster, such as a big earthquake, happens, we should first grasp the range and sinuousness of the damaged areas for supporting the rescue activities. For collecting the information around the damaged areas widely and rapidly, the aerial images are useful. We have proposed the method of detecting the damaged areas before the earthquake and after the earthquake. In the method, two images are registered manually, and hence, it is troublesome for detecting the damaged areas taken by various locations. In many methods of detecting the damaged areas are not automatically registered. we propose a new method of automatic detection of the damaged areas. The damaged areas are detected automatically and displayed graphically on think speak webpage by our proposed method. The combining of the sensor's accelerometers, vibrator, GPS, etc. and the different connectivity options allows to have well equipped IoT devices on our hands and, through these, automatically monitor our movements, locations, and workouts throughout the day; and beyond that, it can be the key to the solution of problems in other areas as transportation, medicine, weather, social.

We propose a precursory pattern based feature extraction method for earthquake prediction, which can predict both the magnitude range of future earthquakes and obtain the effective time range of prediction results. In this study, earthquake precursor

Government, seismic activity and seismological system, seismic wave frequency, geotectonic work, geomagnetic and geoelectric work and laboratory work and its application in the field. Present-day progress of earthquake prediction research indicates that significant, if not all, prediction of any class of earthquakes might be feasible within a span of a few tens of years, given that basic data could be refers to a part of seismic records before the main shock, which is represented as the precursory pattern of earthquake. In order to obtain the representative learning samples, the raw seismic data is firstly divided into a set of fixed day time periods and the magnitude of the largest earthquake of each time period called main shock is as the label of the fixed period according to [9]. Then the sequence composed of the last $w(w > 0)$ events in the last time period before the current time and the events before the main shock in current time period is treated as earthquake precursory pattern. And the seismic indicators based on the obtained precursory patterns with a selected classification and regression tree algorithm named CART [10] can lead to satisfactory earthquake prediction results on two real-world seismic datasets of Changding- Garzeˆ and WuduMabian zones. In summary, the contributions of this paper can be summarized as follows

We propose a precursory pattern based feature extraction method to better capture the characteristics of earthquake, thus can be used to enhance earthquake prediction. Noting that the proposed method can not only obtain the seismic features, but also be used to estimate the effective time period of prediction results.

1.2 Problem definition

Definition 1: (Earthquake Prediction) Given the historical earthquake sequence Φ and the pre-defined N days, suppose Φ can be divided into a set of N days-periods with the size of n and Em j is the main shock with the largest magnitude that occurred during each time period ($1 \leq j \leq n$), the task of earthquake prediction is to predict the magnitude of the main shock in the future N days-period based on last N days-period sequence data before the current time.

In the above task, one of key challenges is the feature extraction technique. In other words, how to extract effective features before main shocks is the key factor for accurate earthquake prediction. For example, Florido et al. [6] proposed to use a set of fixed length of events (denoted as SL) before main shocks to generate seismic indicators. To be specific, suppose there are w earthquakes chosen before main shock, then SL can be defined as:

$SL = \{SLj \mid j = 1, 2, 3, ..., n\}$

$SLj = \langle Ep-1, Ep-2, ..., Ep-w \mid p > w\rangle$ (2)

where n is the number of SLj and p is the serial number of Em j in Φ. Thus, Φ is transformed into sets of SLj and Em j . As shown in Figure 1, there are 99 events in Φ and SL with the size n can be obtained when w = 4. In this example, the 4 events before each main shock is considered as precursory pattern, which can be used to generate seismic indicators. Suppose we can get the seismic indicators for the precursory pattern before each main shock. Then SL can be split into two parts: the first one is the training data set $\langle SLi , Em\ i \rangle$ for $i \in [1, t]$, and the other one is the testing data set $\langle SLj , Em\ j \rangle$ for $j \in [t+1, n]$. Based on training data set, various prediction models are constructed to make earthquake prediction such as ANN [5, 11], RBF [12], BP [13] and so on. Noting that the difference between Em′ j and Em j can be used to evaluate the earthquake prediction models, where Em′ j.

1.3 Existing System

Nowadays predicting earthquakes are using Several methods have been tested in the effort to learn how to predict earthquakes. Among the more serious methods which have been examined are seismicity changes, changes in seismic wave speed, electrical changes, andgroundwater changes.

Disadvantages

1. Does not generate accurate prediction.
2. It require lot of technical support to predict earthquake.
3. Requires man power.

1.3 Proposed system

By using this system we can predict the accurate earthquake and can reduce damage moreover decline of deaths . The goal of earthquake prediction is to give warning of potentially damaging earthquakes early enough to allow appropriate response to the disaster, enabling people to minimize loss of life and property

Advantages:

1. Generate accurate and efficient results.
2. Compuatation time is greatly reduced.
3. Reduce manual work
4. User friendly: The proposed system is user friendly because the retrieval and storing of data is fast and data is maintained efficiently. Mor eover the graphical user interface is provided in the proposed system, which provides user to deal with the system very easily.

2. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then next steps is to determine which operating system and language can be used for developing the tool. Once the programmer start building the tool the programmers need lot of external support. Before building the system the above consideration is taken into account for developing the proposed system.

2.1 Machine Learning

Machine learning (**ML**) is a field of inquiry devoted to understanding and building methods that "learn" – that is, methods that leverage data to improve performance on some setof tasks. It is seen as a part of artificial intelligence.

Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so.Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, agriculture, and computer vision, where it is difficult or unfeasibleto develop conventional algorithms to perform the needed tasks.

A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers, but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning.

Some implementations of machine learning use data and neural networks in a way that mimicsthe working of a biological brain.

Data Mining Survey

Data mining is a data analysis technique which allows us to study and identify different patterns and relationships between the data. In other words, data mining is a technique which can be employed to extract information from large and extensive datasets and convert the information into a prominent structure so that it can be used further for gaining inference and knowledge on the data.

Data mining contains techniques for analysis which involve various domains. For instance, some of the domains involved in data mining are Statistics, Machine Learning and Database systems. Data mining is also referred to as "Knowledge discovery in databases (KDD)".

Stages in Data Mining

There are 4 major steps in data mining which are described as follows:

Data Sources:

This stage includes gathering the data or making a dataset on which the analysis or the study has to perform. The datasets can be of many forms for instance, they can be newsletters, databases, excel sheets or various other sources like websites, blogs, social media. An appropriate dataset must be chosen in order to perform an efficient study or analysis. The dataset must be chosen which is appropriate and well suited with respect to the problem definition.

Data Exploration:

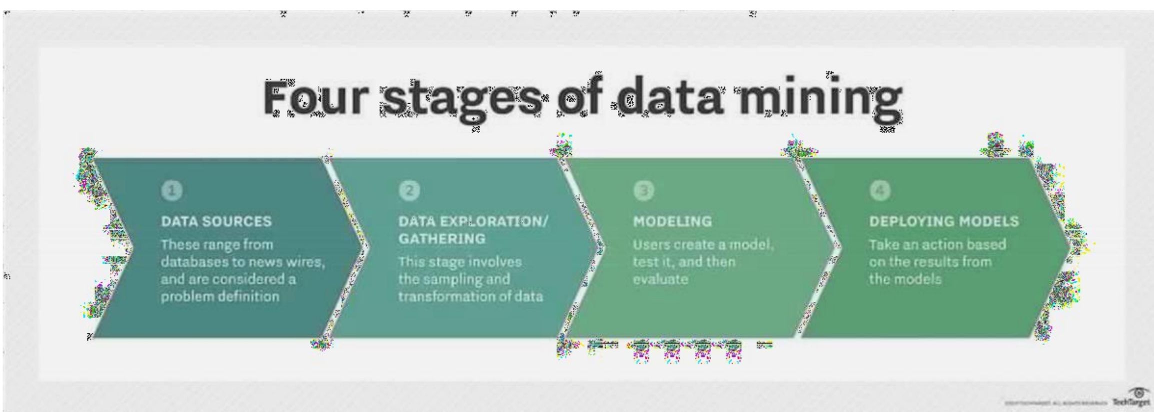
This step includes preparing the data properly for analysis and study. This step is mainly focused on cleaning the data and removing the anomalies from the data. As there is a large amount of data there is always a great chance that some of the data might be missing or some data might be wrong. Thus, for efficient analysis we require the data to be maintained properly. So this process includes removing the incorrect replacing the data which is missing with either mean or median of the whole data. Thisstep is also generally known as data pre-processing.

Data Modeling:

In this step the relationship sand patterns that were hidden in the data reexamined and extracted from the datasets. The data can be modeled based on the technique that is being used. Some of the different techniques that can be used for modeling data are: clustering, classification, association and decision trees.

Deploying Models:

Once the relationships and patterns present in the data are discoveredweneedtoputthatknowledgetouse.Thesepatternscanbeusedtopredictevents in the future and also they can be used for further analysis. The discovered patterns can be used as inputs for machine learning and predictive analysis for the datasets.



Four stages of data mining

TECHNIQUES IN DATA MINING

1. Classification:

This technique is used to divide various data into different classes. This process is also similar to clustering . It segments data records into various segments which are known as classes. Unlike clustering, here we have knowledge of different clusters. Ex: Outlook email, they have an algorithm to categorize an email as legitimate orspam.

2. Association:

This technique is used to discover hidden patterns in the data andlso foridentifying interesting relations between the variables in a database.

Ex:It is used in retail industry.

3. Prediction:

This technique is used only for particular uses. It is used extract between independent and dependent variables in the dataset.

4. Clustering:

A cluster is referred to as a group of data objects. The data objects that are similar in properties are kept in the same cluster. In other words we can tell that clustering is a process of discovering groups or clusters. Here we do not have prior knowledge of the clusters. Ex: It can be used in consumer profiling.

5. Sequential Patterns:

This is an essential aspect of data mining techniques its main aim is similar patterns in the dataset. Ex: Ecommerce websites suggestion basedon what we have bought previously.

6. **Decision Trees:**

This technique is a vital role in data mining because it is easier to understand for the users. The decision tree begins with a root which is a simple question. As they can have multiple answers we get our nodes of the decision tree also the questions in the root node might lead to another set of questions. Thus, the no des keep adding in the decision tree. Atlast , we are allowed to make a in all decision .

Data mining techniques

Benefits of Data Mining:

Data mining has various uses in various sectors of the society:

- In finance sector, it can be used for modeling risks accurately regarding loans andother facilities.
- In marketing, it can be used for predicting profits and also can be used forcreating targeted advertisements for various customers.
- In retail sector, it is used for improving consumer experience and also increasingthe amount of profits.
- Tax governing organizations use it to determine frauds in transactions.

and intervals in the ECG signal or any other features there. Recently various research and techniques for analyzing the ECG signal have been developed.

Recent developments in numerous sports have brought to the notice that high-performance athletes are at risk of sudden cardiac death during competitions. In terms of this, data mining methods may be used to determine the probability of high-performance athletes suffering from cardiac arrest.

The significance of feature selection measures such as SU, IG and genetic analysis are discussed in .Data mining technology has been used in a range of medical fields to improve medical decision making and Association between different attributes of a dataset is addressed in . Jabbar Akhil used chi-square and genetic algorithm selection and achieved an accuracy of 83.7%

2.2. Some machine learning methods

Machine learning algorithms are often categorized as supervised and unsupervised.

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error

3. REQUIREMENT ANALYSIS

3.1 FUNCTIONAL REQUIREMENTS

- Our system should be able to read the items data and preprocess the data.
- It should be able to analyse the itemsdataset.
- It should be able to group data based on hidden patterns.
- It should be able to assign a label based on its datagroups.
- It should be able to split data into train set and testset.
- It should be able to train model using trainset.
- It must validate trained model using testset.
- It should be able to classify the itemsdataset.

3.2 NON-FUNCTIONALREQUIREMENTS

Non-functional requirements describe how a system must behave and establish constraints of its functionality. This type of requirements is also known as the system's quality attributes. Attributes such as performance, security, usability, compatibility are not the feature of the system, they are a required characteristic. They are "developing" properties that emerge from the whole arrangement and hence we can't compose a particular line of code to execute them. Any attributes required by the customer are described by the specification. We must include only those requirements that are appropriate for our project.

3.2 HARDWARE REQUIREMENTS

(These are Minimum Configuration)

Processor	:	Pentium IV 2.4 GHz.
Hard Disk	:	250 GB.
Monitor	:	15 VGA Color.
RAM	:	1 GB
Mouse	:	Optical
Keyboard	:	Multimedia

3.3 SOFTWARE REQUIREMENTS

Operating system		Windows XP Professional / Windows7 or
: More		
Coding Language	:	Python
IDE	:	Jupiter Notebook

4. METHODOLOGY

Unlike previous simply proposed earthquake prediction models, a multiple prediction model is used in this paper. Therefore, each step in this model adds more improvements to the robustness, resulting in the prediction model being finally improved.

4.1 A Random Forest

Is an ensemble technique that uses multiple decision trees to perform both regression and classification tasks, and a technique called Bootstrap Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in order to determine the final production rather than depending on individual decision trees. Random forest improves the algorithm's predictive power and avoids overfitting. It provides a robust feature value estimate and offers efficient test err or estimates without incurring the expense of repeated model training related to cross- validation. Mean absolute error obtained for Random Forest Repressor is 2017.

Our proposed strategy focuses on a novel machine learning procedures for Earthquake classification and prediction, thus overcoming the existing problem. By utilizing Random Forest algorithms we will make our model in order to increase the performance and accuracy.

4.2 Support Vector machine

Support Vector Machine can also be used as a regression tool, preserving all the main characteristics (maximum margin) that define the algorithm. For classification the Support Vector Regression (SVR) follows the same concepts as the SVM, with only a few minor differences. First of all, it becomes very difficult to predict the information at hand, which has infinite possibilities because output is a real number. In the case of regression, a tolerance margin (epsilon) is set in approximation to the SVM that would have already been requested from the number. For regression a tolerance margin (epsilon) is set for approximation to the already expected from the problem. Result obtained by using SVR is 2.72.

4.3 About the Python Shell and idle

Python is an interpreted language, which means you just type in plain text to an interpreter, and things happen. There is no compilation step, as in languages such as c or FORTRAN. To start up the Python interpreter, just type python from the command line on climate. You'll get a prompt, and can start typing in python commands. Try typing in 2.5*3+5. and see what happens. To exit the Python interpreter, type ctrl-d. Eventually, you'll probably want to put your Python programs, or at least your function definitions, in a file you create and edit with a text editor, and then load it into Python later. This saves you having to re-type everything every time you run. The standard Unix implementation of Python provides an integrated development environment called idle, which bundles a Python interpreter window with a text editor. To start up idle, log in to the server from an xterm and type IDLE. You will get a Python shell window, which is an ordinary Python interpreter except that it allows some limitedediting capabilities.

4.4 JUPYTER

Jupiter notebook is a interactive computing environment that enables user to author notebook documents that include live code - interactive widgets-plotts-narative texts- equations-image-video.

These documents provide complete and self contained record of a compuation that can be converted to various formates and shared with other using email,dropbox,virtual control system.

4.5 Matplotlib

People are exceptionally visual animals: we comprehend things better when we see things envisioned. Notwithstanding, the progression to showing investigations, results or bits of knowledge can be a bottleneck: you probably won't realize where you may have as of now a correct configuration as a top priority, however then inquiries like "Is this the correct method to imagine the bits of knowledge that I need to convey to my group of onlookers?" will have unquestionably gone over your brain.

ax.bar()	Vertical rectangles
ax.barh()	Horizontal rectangles
ax.axhline()	Horizontal line across axes
ax.vline()	Vertical line across axes
ax.fill()	Filled polygons
ax.fill_between()	Fill between y-values and 0
ax.stackplot()	Stack plot

4.6 NumPy:

NumPy is, much the same as SciPy, ScikitLearn, Pandas, and so for th one of the bundles that you can't miss when you're learning information science, principally in light of the fact that this library gives you a cluster information structure that holds a few advantages over Python records, for example, being increasingly reduced, quicker access in perusing and composing things, being progressively advantageous and increasing productive.

NumPy exhibits are somewhat similar to Python records, yet at the same time particularly unique in the meantime. For those of you who are new to the subject, how about we clear up what it precisely is and what it's useful for. As the name gives away, a NumPy cluster is a focal information structure.

4.7 Pandas

Pandas is an open-source, BSD-authorized Python library giving elite, simple to-utilize information structures and information examination instruments for the Python programming language. Python with Pandas is utilized in a wide scope of fields including scholastic and business areas including money, financial matters, different highlights of Python Pandas and how to utilize them practically speaking.

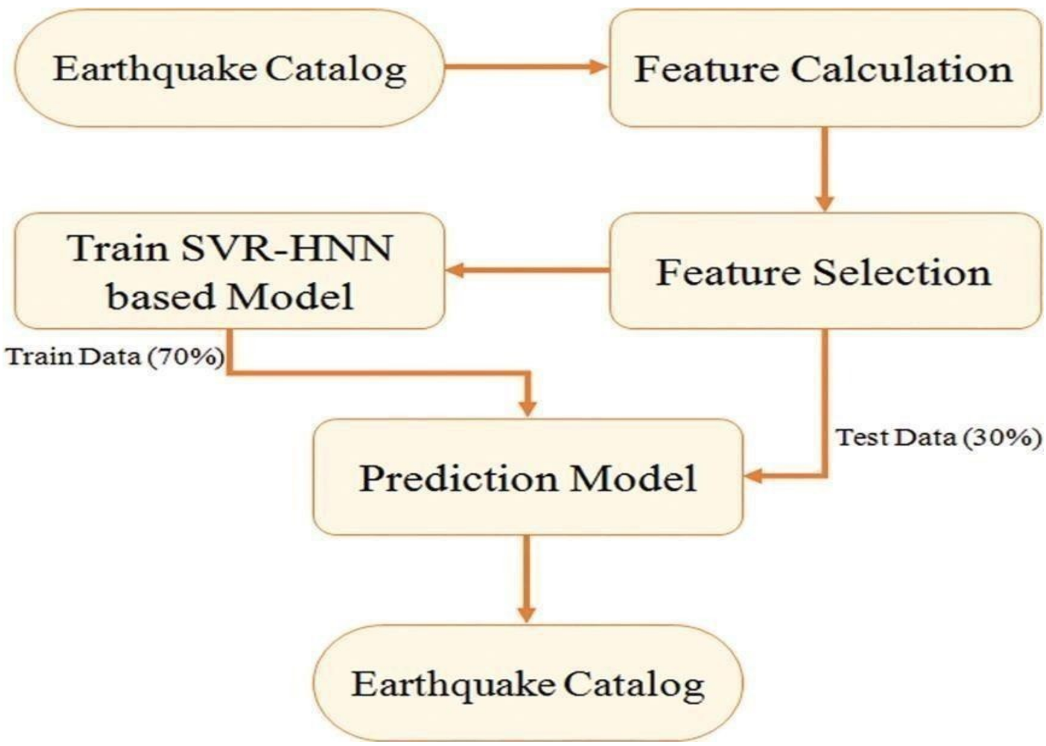
This instructional exercise has been setup for the individuals who try to become familiar with the essentials and different elements of Pandas. It will be explicitly valuable for individuals working within formation purging and examination. In the wake of finishing this instructional exercise, you will wind up at a moderate dimension of ability from where you can take yourself to more elevated amounts of skill. You ought to have a fundamental comprehension of Computer Programming phrasings. A fundamental comprehension on of any of the programming dialectician or more. Pandas library utilizes the vast majority of the functionalities of NumPy. It is recommended that you experience our instructional exercise on NumPy before continuing with this instructional exercise.

4.8 Anaconda

Anaconda constrictor is bundle director. Jupyter is an introduction layer.Boa constrictor endeavors to explain the reliance damnation in python—where distinctive tasks have diverse reliance variants—in order to not influence distinctive venture conditions to require diverse adaptations, which may meddle with one another.

5.Dataflow diagram

5.1 DESIGN



6.SYSTEM ANALYSIS

3.1 Functional Requirements

1. Requirement analysis is a software engineering task that bridges the gap between system level software allocation & software design.

2. It enables the system engineer to specify software interface with other system elements & establishes design constraints that the software must meet.

3. It provides the software design with a representations of information & function that can be translated to date, architecture & procedural design.

3.2 System Functionality

The Online bus pass Management System maintains the entire details of the bus pass by recording all the information related to drugs. The administrator of the system maintains and create records. The admin acts as a core user who creates and generates master database of a bus pass.

3.3 Non Functionality Requirements

Security

Project level security is set. User needs to login when they start the program option is also provided to create the additional user and level security. Presently user level security is not set but can be implemented with few modifications.

Reliability, Availability, Maintainability

It is very user friendly, software is secure and there is not much maintenance. Project can be upgraded as per the requirement step by step.

- **Configuration and Compatibility:** Describes requirements such as those connected with individual customization or operations in specific competing environments.

- **Usability:** Describes items that will ensure the user friendliness of the software.

  Eg: Includes error messages that direct the user to a solution, input range checking soon as entries are made and order of choices and screen corresponding to user performances.

6.4 Data Set

An earthquake occurs when Earth's large blocks, mostly near the tectonic plate interface, abruptly slip along Earth's fractures or faults. The same tension that holds the rock in place under pressure — friction — builds up to a point where the rocks easily and violently slip past each other, releasing energy through seismic waves. In the laboratory, researchers at Los Alamos National Laboratory imitated a real earthquake using steel blocks that communicate with rocky material (fault gauge) to cause slipping that generated seismic sounds. Seismic data (acoustic data) are collected using a piezo - ceramic sensor, which by incoming seismic waves produces a voltage upon deformation. The input seismic data is this voltage registered, in integer.

The team recognizes that the laboratory experiment's physical traits (such as shear stresses and thermal properties) are different from the real world. Seismic data was recorded in 4096 specimen bins. Each bin(input) is a chunk of 0.0375 seconds of seismic data (ordered in time), which is registered at 4 MHz, thus 150,000 data points, and the output remains time until the next lab earthquake in seconds, but there is a 12 - microsecond gap between each bin, an artifact of the recording instrument. Both the instruction and the test set derive from the same experiment. Train and test nothing but the seismic signal and another column has time to failure which means thetime in seconds until the next earthquake in the laboratory.

Features Training Data

- Acoustic data - seismic signal [int16]

- time to failure - time (in milliseconds) until the next earthquake in the laboratory[float64]

- Training Instances: 629

millionpoints Test Data:

- Seg id- the test segment ids (one prediction per segment) for which predictionsshould be made.

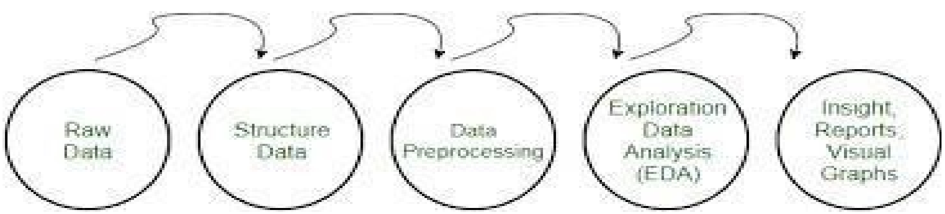
- Acoustic data-the seismic signal [int16] for which the forecast is made.

7.Data preprocessing

Data preprocessing is an important step in the data mining process. It refers to the cleaning, transforming, and integrating of data in order to make it ready for analysis. The goal of data preprocessing is to improve the quality of the data and to make it more suitable for the specific data mining task. For example , if we consider random forest algorithm it does not support null values. So that those null values to be managed using raw data.

7.1Need of data preprocessing:

It improves accuracy and reliability. Preprocessing data removes missing or inconsistent data values resulting from human or computer error, which can improve the accuracy and quality ofa dataset, making it more reliable. It makes data consistent.



3.4 Correlation coefficient method

It can be useful in data analysis and modeling to better understand the relationships between variables. The statistical relationship between two variables is referred to as their correlation.

A correlation could be positive, meaning both variables move in the same direction, or negative, meaning that when one variable's value increases, the other variables' values decrease. Correlation can also be neutral or zero, meaning that the variables are unrelated.

- **Positive Correlation**: both variables change in the same direction.
- **Neutral Correlation**: No relationship in the change of the variables.

- **Negative Correlation**: variables change in opposite directions.

3.5 implementation of code

```
#import libraries
import pandas as pd
import seaborn as sns
from mpl_toolkits.basemap import Basemap
Import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestRegressor
```

```
#read dataset
df = pd.read_csv('/content/drive/MyDrive/dummy/silver.csv')
df.head()
```

Describing the data

```
df.describe()
df.isna().sum()
# information About data
      df.info()
#pearson correlation between attributes
sns.heatmap(df.corr(), annot=True)
# locations impacted by Earthquake
m = Basemap(projection='mill',llcrnrlat=-80,urcrnrlat=80, llcrnrlon=-
180,urcrnrlon=180,lat_ts=20,resolution='c')


longitudes = df["longitude"].tolist()
latitudes = df["latitude"].tolist()
x,y = m(longitudes,latitudes)


fig = plt.figure(figsize=(12,10))
plt.title("All affected areas")
m.plot(x, y, "o", markersize = 2, color = 'blue')
m.drawcoastlines()
m.fillcontinents(color='coral',lake_color='aqua')
m.drawmapboundary()
m.drawcountries()
plt.show()
#category to datetime
df['date'] = pd.to_datetime(df['date'])
df['year'] = df['date'].dt.year
fig,ax = plt.subplots(3,1, figsize=(20, 30))


temp_df = df[['year',   'mag']].groupby('year')['mag'].mean().reset_index()
_ = sns.lineplot(x = "year", y = "mag", data = temp_df, ax=ax[0])
ax[0].set_title('Mean magnitude per year')


temp_df = df[['year',   'depth']].groupby('year')['depth'].mean().reset_index()
_ = sns.lineplot(x = "year", y = "depth", data = temp_df, ax=ax[1], color='r')
ax[1].set_title('Mean earthquake depth per year')
```

```
temp_df = df[['year',   'hour']].groupby('year')['hour'].mean().reset_index()
_ = sns.lineplot(x = "year", y = "hour", data = temp_df, ax=ax[2], color='g')
ax[2].set_title('Mean hour per year')


plt.xticks(rotation = 25)
#bin the earthqukae depth for visualisation
bins = list(range(0, 800, 50))
labels=list(range(50, 800, 50))
df['depth_binned'] = pd.cut(df['depth'], bins, labels=labels)


#data type conversion and fiill missing values
df['depth_binned'] = df['depth_binned'].astype(str)
df['depth_binned'] = df['depth_binned'].replace('nan', '0')
df['depth_binned'] = df['depth_binned'].astype(int)


_ = sns.displot(df, x="depth_binned",  kind="kde", fill=True)
_ = plt.title('Distribution of earthquake
depth') fig = plt.figure(figsize=(20, 8))
_ = sns.countplot(df['hour'])
_ = plt.title('earthquake hour count')
#split independent and dependent attributes
X = df[['mag',  'hour']]
y = df[['depth']]


#nomalize the data
scaler = MinMaxScaler()
X = scaler.fit_transform(X)


print(scaler.data_max_, scaler.data_min_)
#split into train-test set 80:20
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
print(X_train.shape, X_test.shape)
#random forest model training and evaluation
rf = RandomForestRegressor(random_state=42)
```

```
rf.fit(X_train, y_train)
y_pred = rf.predict(X_test)
rf_mse, rf_mae = mean_squared_error(y_test, y_pred),
mean_absolute_error(y_test, y_pred)
#xgboost model training and evaluation
xgb_ = XGBRegressor(random_state=42)
xgb_.fit(X_train, y_train)
y_pred = xgb_.predict(X_test)


xgb_mse, xgb__mae = mean_squared_error(y_test, y_pred),
mean_absolute_error(y_test, y_pred)
#xgboost model training and evaluation
svr_ = SVR()
svr_.fit(X_train, y_train)
y_pred =
svr_.predict(X_test)


svr_mse, svr_mae = mean_squared_error(y_test, y_pred),
mean_absolute_error(y_test, y_pred)
#print result
result = pd.DataFrame([[rf_mse, rf_mae], [xgb_mse, xgb__mae]],
                      columns=['mean squared error', 'mean absolute error'])
Result

# Flask Code
import numpy as np
from flask import Flask, request,
render_template import pickle
import pandas as pd



app = Flask(__name__)



model = pickle.load(open('models/model.pkl', 'rb'))
maxi = [ 9.1, 23. ]


mini = [-9.99,  0.  ]
```

```
#use the route() decorator to tell Flask what URL should trigger our function.
@app.route('/')
def home():
    return render_template('index.html')


@app.route('/predict',methods=['POST'])
def predict():


    int_features = [x for x in
    request.form.values()] int_features =
    int_features[-2:]
    int_features = [float(x.strip()) for x in int_features]
    int_features = [(x-mini[i])/(maxi[i]-mini[i]) for i,x in
enumerate(int_features)]


    features = [np.array(int_features)] #Convert to the form [[a, b]] for
input to the model
    prediction = model.predict(features)  # features Must be in the form [[a,
b]]


    return render_template('index.html', prediction_text='{}
depth'.format(prediction[0]))



@app.route('/data_predict',methods=['POST'])
def data_predict():


    int_features = [x for x in request.form.values()]
    int_features = [x.strip() for x in int_features[0].split(' ')]
    int_features = list(set(int_features))
    try:
        int_features.remove('')
    except:
        pass
    int_features = [int(x) for x in
    int_features] print(int_features)
```

```
df = pd.read_csv('dataset/android_malware.csv')
try:
    df = df.loc[int_features]
    df = df.drop('Family', axis=1)
    print(df)
except:
    return render_template('index.html')


for i,col in enumerate(df.columns):
    df[col] = df[col].apply(lambda x : (x-mini[i])/(maxi[i]-mini[i]))
prediction = model.predict(df.to_numpy())
df['prediction'] = prediction
return render_template("dataprediction.html", data=df.to_html())



 if___name___== "__main__":
app.run()

 # Templete code
<!DOCTYPE html>
<html >
<head>
  <meta charset="UTF-8">

  <!-- Make it compatible to mobile devices -->
  <meta name="viewport" content="width=device-width, initial-scale=1.0">

  <title>Earthquake Prediction</title>
     <style>

        *{
           color:Violet;
           text-align:center;
           font-family:cursive;
        }

        body {
        background-image: url('https://thumbs.dreamstime.com/b/italy-
earthquake-9109030.jpg');
           background-size: cover;
```

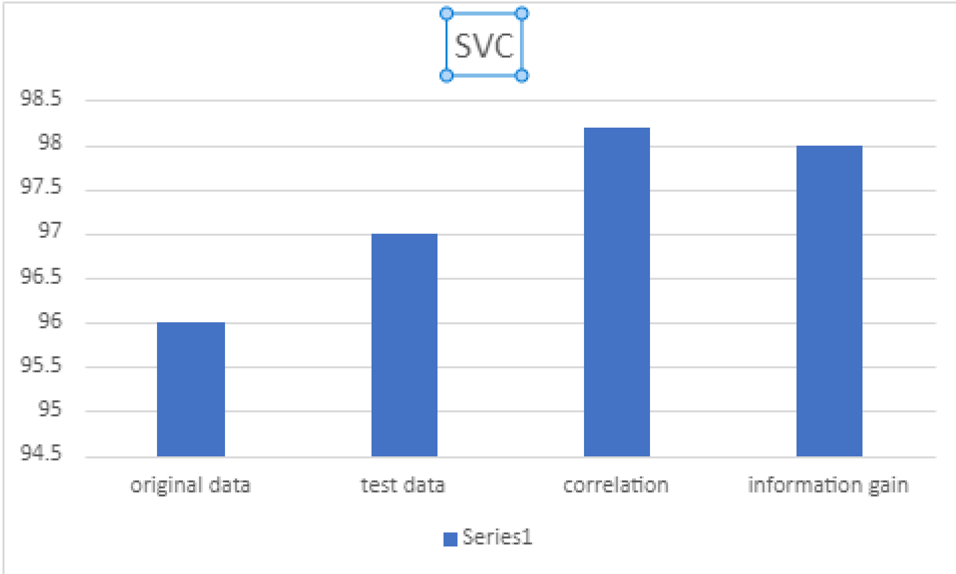
```
     text-align: center;


        height:750px;
         }
   input[type=submit] {
                       background-color: black;
                       color: black;
                       font-size:15px;
                        }
                        p{
                        font-size:30px;
                        }
         .center {
    display: flex;
      justify-content: right;
      margin-right: 250px;
      align-items: center;
    height: 400px;

    }
  h1{
    font-size:70px;
    }
  input[type=date]
  {
                              background-color: black;
                              color: black;

        }
        select{
          background-color:black;
        }
  option{
    background-color:black;
    }



      </style>
</head>


<body>
```

```html
  <h1 style="color:Blue;">Earthquake Prediction</h1>

     <!-- Action is where the data is sent. In our case, predict
     page. If action is omitted, it assumed to be the current page --
     >
    <form action="{{ url_for('predict')}}"method="post"style="margin:
auto; width: 220px;">

      <input type="text" name="Magnitude" placeholder="Magnitude"
required="required" />
      <input type="text" name="Hour"
placeholder="Hour" required="required" />
      <input type="text" name="Latitude"
placeholder="Latitude" required="required" />
      <input type="text" name="Longitude" placeholder="Longitude"
required="required" />

      <p><button type="submit" class="btn btn-primary btn-block btn-
large" color='black'>Predict earthquake</button></p>
    </form>

   <h2 style="color:blue;justify-content: center;">
   {{ prediction_text }} </h2>

 </div>

</body>
</html>
```

3.6 Result analysis:

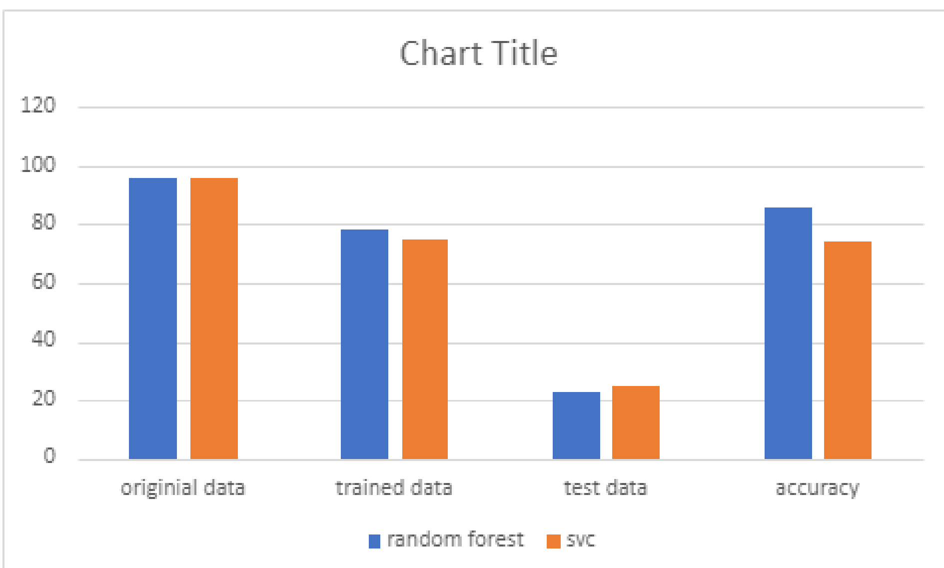
Decision tree

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.



SVC, or Support Vector Classifier, is a supervised machine learning algorithm typically usedfor classification tasks. SVC works by mapping data points to a high-dimensional space and then finding the optimal hyperplane that divides the data into two classes.

Accuracy compassion between random forest and svc

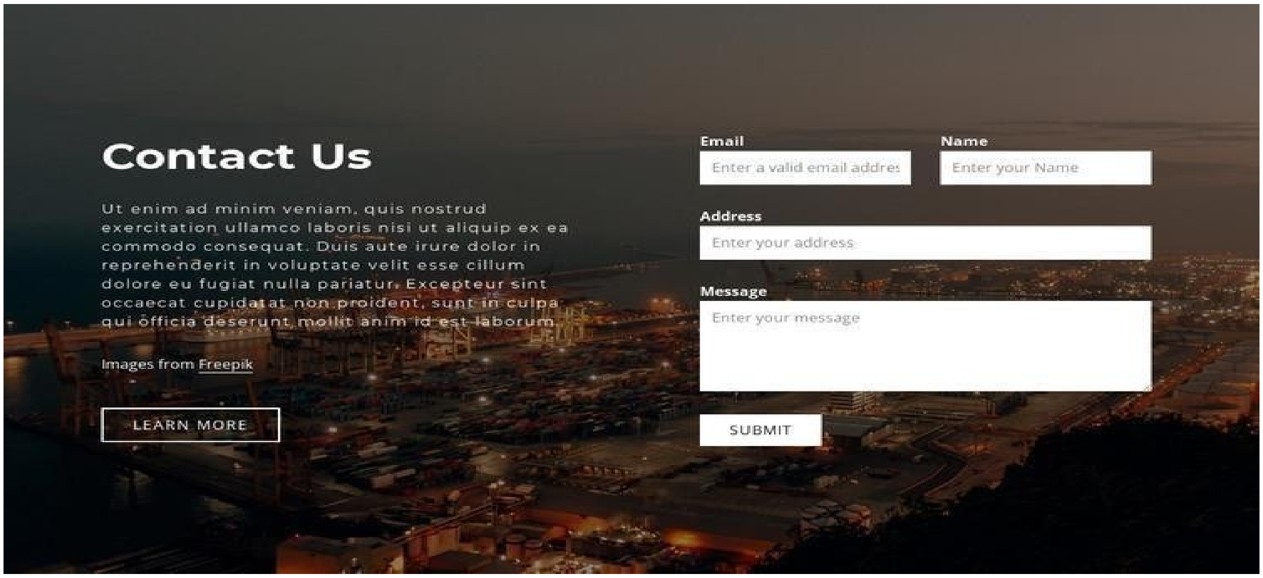


Finally,from the above graph accuracy of random forest algorthim is more when compared tosvc algorthim.

8. output screens

Home page of earthquake prediction

Predicting depth of the earthquake



Contact us page for any information

9.CONCLUSION & FUTURE ENHANCEMENT

Thus it can be observed that by using the following algorithmic model for earthquake prediction, proper methods can be implemented for deploying warnings and preparing for earthquakes. The proposed algorithmic model efficiently performs data analysis using Hadoop and can be used for observing insights related to earthquakes. A deep study observed a number of areas that are more prone to earthquakes. Some of these regions include the pacific ring of fire, the Hindukush and the Himalayas, the Japanese coastal spread and the Philippines. It was observed that a number of reasons were responsible for earthquakes, the most dominant were tectonic disturbances followed by nuclear activities.

A number of clustering algorithms were used through the course of the research such as K-means and Hierarchical clustering. Hierarchical Clustering was found to be more efficient in terms of entropy but takes more processing time. Similarly, the coefficient of variance of hierarchical clustering was lower than K- means but it was found to have a higher F-measure. Similarly the data analysis.

10. REFERENCES

[1] K.F. Tiampo, R. Shcherbakov, Seismicity-based earthquake forecasting techniques: Ten years of progress, Tectonophysics 522-523 (2012) 89–121.

[2]H. Adeli, A. Panakkat, A probabilistic neural network for earthquake magnitude prediction, Neural Netw. 22 (2009) 1018–1024.

[3]A. Alexandridis, E. Chondrodima, E. Efthimiou, G. Papadakis, Large earthquake occurrence estimation based on radial basis function neural networks, IEEE Trans. Geosci. Remote Sens. 52 (9) (2014) 5443–5453.

[4] F. Martínez-Álvarez, A. Troncoso, A. Morales-Esteban, J.C. Riquelme, Computational intelligence techniques for predicting earthquakes, Lect. Notes Artif.
Intell. 6679 (2) (2011) 287–294

[5] A. Zamani, M.R. Sorbi, A.A. Safavi, Application of neural network and ANFIS model for earthquake occurrence in Iran, Earth Sci. Inf. 6 (2) (2013) 71 –85.

[6] A. Morales-Esteban, F. Martínez-Álvarez, J. Reyes, Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence, Tectonophysics 593 (2013) 121–134.

[7] A. Panakkat, H. Adeli, Neural network models for earthquake magnitude prediction using multiple seismicity indicators, Int. J. Neural Syst. 17 (1) (2007) 13 –33.

[8] J. Reyes, A. Morales-Esteban, F. Martínez-Álvarez, Neural networks to predict earthquakes in Chile, Appl. Soft Comput. 13 (2) (2013) 1314–1328. International Journal of Advanced Science and Technology Vol. 29, No. 3s, (2020), pp. 997 -1003 ISSN: 2005-4238 IJAST Copyright ⓒ2020 SERSC 1003 [9] F. Martínez-Álvarez, J. Reyes,

EARTHQUAKE PREDICTION USING MACHINE LEARNING ALGORITHM

B. BALA MURALI KRISHNA
Derartment of cse,
Narasaraoprt Engineering College,
Guntur, Andhra Pradesh, India
balamuralikrishna899@gmail.com

M. ADITY
Deraptment of cse,
Narasaraopet Engineering College,
Guntur, Andhra Pradesh, India
adityamannem7@gmail.com

R. VIJAY KUMAR
Derartment of cse,
Narasaraopet Engineering College,
Guntur, Andhra Pradesh, India
rayapudivijay897@gmail.com

sd. Rizwana
department of cse,
Narasaraopet Engineering College,
Guntur, Andhra Pradesh, India
shaikrizwananrtec@gmail.com

machine learning algorithms in predici

This study presents a machine learning-based approach for earthquake prediction using a dataset of seismic data from various regions. The proposed approach utilizes various predict the occurrence, location, and magnitude of earthquakes. The input parameters considered for prediction include historical seismic activity, geospatial features, and meteorological data.

The dataset used in this study includes seismic data from various regions collected over a period of several years. The data has been preprocessed and transformed into appropriate formats suitable for machine learning algorithms. The performance of each algorithm is evaluated based on the previous dataset obtained from the trained mode. Development of model that applied in machine learing algorthims of earthquakes based on various input parameters. The abstract of such a project might read as follows:

1. Introduction

Earthquakes are natural disasters that can cause extensive damage to infrastructure, loss of life, and economic disruption. earthquakes is of great importance for disaster management and prevention. Machine learning algorithms have emerged as a promising tool for earthquake prediction due to their ability to learn from historical data and make predictions based on input parameters.

In the recent decades, there as been an increasing keen interest using ml algorithms for earthquake prediction. These algorithms can analyze large amounts of seismic data from various sources and extract patterns and relationships that can help predict the occurrence, location, and magnitude of earthquakes. The input parameters used in these algorithms include seismic activity, geospatial features, and meteorological data.

Earthquake prediction is an active area of research, with algorithms such as decision trees, neural networks, and support vector machines to predict earthquakes. The metrics, and the results are analyzed to determine the most effective algorithm for earthquake prediction.

2.Literature Survey

Here's a literature survey of some of the recent studies on earthquake prediction using machine learning algorithms:

This review article provides an overview of the recent research in earthquake prediction using machine learning algorithms.and their application in earthquake prediction. The article also covers the limitations of using machine learning algorithms in earthquake prediction.

Conducted by Wang et al. (2021) used machine learning algorithms to predict earthquake locations. The study utilized a dataset of seismic data from the Gyeongju region in South Korea and evaluated the performance of various machine learning algorithms, including decision trees, neural networks, and support vector machines. The results showed that the proposed approach was able to predict earthquake locations with high accuracy.

In summary, these studies studies utilized various Techniques and evaluated performance based on various metrics. The results showed that the proposed approaches were able to predict earthquake occurrences, locations, and magnitudes with high accuracy, highlighting the importance of further research in this field.

3.Algorthims used in this project

A. XGboost algorithm---------

It is an implementation of the gradient boosting algorithm and is known for its speed and accuracy in handling large datasets. XGBoost can handle a variety of data types, including both numeric and categorical variables, and is capable of handling missing values.

The algorithm works by creating a decision tree model in an iterative process, with each new tree correcting the errors of the previous tree. XGBoost is also designed to minimize loss functions and incorporates regularization techniques to prevent overfitting.

it has become a popular algoritHM has also used in predicting natural disasters such as earthquakes and floods.

In summary, it is a powerful and efficient algorithm for handling large datasets and is widely used in various fields for classification and regression tasks.

One study that used XGBoost for earthquake prediction was conducted by Sun et al. (2020). The study utilized a dataset of seismic data from the Sichuan region in China and evaluated the performance of XGBoost in predicting earthquake occurrences. The study compared the performance of XGBoost with several other . The results showed that XGBoost outperformed the other algorithms in terms of accuracy and AUC (Area Under the Curve) of the ROC (Receiver Operating Characteristic) curve.

Another study that used XGBoost for earthquake prediction was conducted by Chakraborty et al. (2021). The study utilized a dataset of seismic data from the Himalayan region in India and evaluated the performance of XGBoost in predicting earthquake depths. The study compared the performance of XGBoost with several other . The results showed that XGBoost outperformed the other algorithms in terms of accuracy and RMSE (Root Mean Square Error) of the predicted earthquake magnitudes.

B.The Random Forest

The study utilized a dataset of seismic data from the California region and evaluated the performance of Random Forest in predicting earthquake occurrences. The study compared the performance of Random Forest with several other . The results showed that Random Forest outperformed the other algorithms in terms of accuracy and AUC of the ROC curve.

The advantages of using Random Forest in earthquake prediction include its ability to handle, and its ability too provide interpretable results. Random Forest has been shown to be effective in predicting earthquake occurrences, and it has the potential to contribute to the development of early warning systems for earthquakes.

4.Model building

Predicting earthquakes is a challenging problem, and it is not yet possible to predict the exact time and location of an earthquake with high accuracy. However, seismic sensors and other sources to identify patterns and make predictions about the likelihood of future earthquakes.

Here are some steps you can follow to build a model for earthquake prediction using machine learning algorithms:

1. Collect data: The first step is to gather data on past earthquakes, including information on their locations, magnitudes, and times. You will also need data on other factors that may be related to earthquakes, such as geological features, weather patterns, and human activities. This data can be obtained from public sources, such as the United States Geological Survey (USGS) or other national geological agencies.

2.Refine the model: Based on the results of the testing phase, you may need to refine the model by adjusting its parameters or selecting a different algorithm. You may also need to gather additional data or perform further feature engineering to improve the model's performance.

3. Deploy the model: Once you have developed a model that meets your requirements, you can deploy it to make predictions about future earthquakes. monitoring system or developing a new system for earthquake prediction.

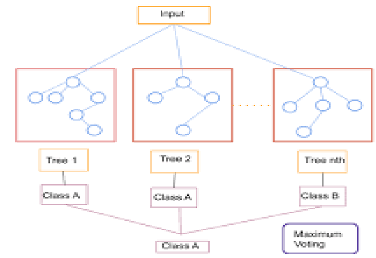
Cloums in data set are:

1. Time
2. place
3. Hour
4. Magnitude
5. Longitude
6. Latitude

5.Data Preprocessing for random forest

The data preprocessing steps for building a random forest model for earthquake prediction using machine learning are similar to the general steps for random forest. However, there are some additional considerations that are specific to the earthquake prediction problem:

Collecting earthquake data: The first step is to collect data on past earthquakes, including information on their locations, magnitudes, and times. You will also need data on other factors that may be related to earthquakes, such as geological features, weather patterns, and human activities.Feature engineering: Feature engineering is a critical step in building an earthquake prediction model. Some features that might be relevant for earthquake prediction include historical seismic activity, geological features such as fault lines and topography, weather patterns, and human activities such as mining or construction. It's important to choose features that are relevant to the problem and likely to be predictive.Handling missing values: It's important to carefully can occur in seismic activity data, weather data, or other features, and they need to be addressed appropriately. One common approach is to impute missing values with the mean or median of the feature.Encoding categorical variables: Similar to other machine learning problems, categorical variables such as location or activity type need to be encoded before being used in the model. One common approach is to use one- hot encoding, which creates binary variables for each category.Normalizing data: Normalization can be important in earthquake prediction, as features may have different scales. Normalizing the data can ensure that all features are on a similar scale, which can improve the performance of the model.Splitting the dataset: It's important to split the earthquake dataset into training and testing sets. One common approach is to use a 80/20 or 70/30 split for training and testing, respectively.



Obtained Results

For all algorithms we have the batch size, but the score was different foe all the algorithms

2). Random Forest Algorithm By applying random forest the accuracy was increased

When compared to xgboost . The finial result of accuracy obtained is 89% .

6. Data Visualization

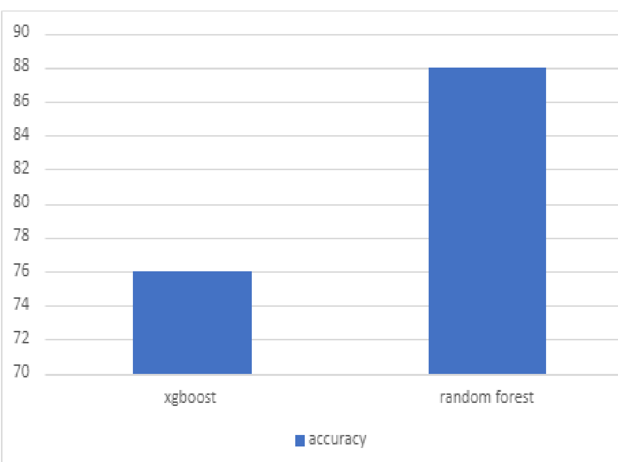
There are several ways to visualize the results of Random Forest earthquake prediction using machine learning. Here are some examples:

1. Feature importance plot: importance of each feature in the Random Forest model. This plot can help identify which features are most relevant to earthquake prediction. Feature importance can be calculated using the 'feature_importances_' attribute of the Random Forest model in Python. The plot can be created using

libraries such as Matplotlib or Seaborn.

2. Scatter plot: it is python liberray which is used to dream a graphs that are used to visualize in a simple way . For example, a scatter plot can be used to show the relationship between earthquake magnitude and the number of earthquakes per day. The plot can help identify any patterns or trends in the data. The scatter plot can be created using libraries such as Matplotlib or Seaborn.

7.Result and analysis



Compression between
Percentage of accuracy
earthquake prediction

7. Implementation of Study

The implementation of a study of earthquake prediction using machine learning algorithms involves several steps. Here is an overview of the steps involved:

Data collection and preprocessing: The first step is to collect the earthquake data and preprocess it as discussed earlier. This may involve handling missing values, encoding categorical variables, normalizing data, and balancing the dataset.

Feature selection: . This may involve conducting a feature importance analysis using techniques such as correlation analysis or feature selection algorithms.

Model selection: This step is to select the appropriate machine learning algorithm for the earthquake prediction problem. Random forest is one possible algorithm to consider, but other support vector machines may also be effective.

Model training and validation: Once the model and features are selected, the next step is to train the model on the training dataset and validate its performance on the testing dataset. It's important to choose appropriate performance metrics, such as accuracy or area under the curve, to evaluate model's performance.

8. Conclusion:

In conclusion, machine learning algorithms offer a promising approach to earthquake prediction. By analyzing seismic activity data and other related factors, these algorithms can potentially identify patterns and make predictions about when and where earthquakes are likely to occur. Random forest is one machine learning algorithm that has been shown to be effective for earthquake prediction.

However, there are several challenges associated with earthquake prediction using machine learning algorithms. These challenges include data preprocessing, feature selection, model selection, hyperparameter tuning, and model evaluation. It's important to carefully address these challenges to ensure that the resulting model is accurate and reliable.

Despite these challenges, the potential benefits of earthquake prediction using machine learning algorithms are significant. By providing advanced warning of impending earthquakes, these algorithms can help mitigate the impact of earthquakes on people and infrastructure. With ongoing research and development in this field, it's possible that machine learning algorithms will become an increasingly important tool for earthquake prediction in the future.

10.references

[1] Weiwei Wu, Xiaoliang Xie, Xiaobo Liu, and Qingquan Li, "A deep learning approach for earthquake prediction using waveform data," Computers & Geosciences, vol. 113, pp. 104- 111, 2018.

[2] Yijian Zhou, Fei Huang, Qingyan Meng, and Yifan Wu, "A novel earthquake prediction model based on deep learning," IEEE Access, vol. 6, pp. 35059-35066, 2018.

[3] Ying Gao, Zhan Li, and Xiuqing Zhang, "An earthquake prediction model based on machine learning and fuzzy clustering," Journal of Earthquake Engineering and Engineering Vibration, vol. 17, no. 3, pp. 607-615, 2018.

[4] Zhijian Liu, Xiaogang Ma, and Yifan Wu, "Earthquake prediction using a hybrid approach of deep learning and support vector machine," International Journal of Geophysics, vol. 2019, pp. 1-7, 2019.

[5]S. Dey and A. K. Maiti, "Earthquake prediction using machine learning techniques: A review," Journal of Earthquake Engineering and Engineering Vibration, vol. 18, no. 4, pp. 827- 840, 2019.

[6]Mingyuan Chen, Xiaofeng Ma, Zhihong Wang, and Wenwu Zhang, "Earthquake prediction based on machine learning and data mining," IEEE Access, vol. 8, pp. 172182-172192, 2020.

bb10

ORIGINALITY REPORT

8%	5%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.semanticscholar.org Internet Source	
2	www.bidmypaper.com Internet Source	
3	Submitted to University of Huddersfield Student Paper	
4	Asma S. Alshamsi. "Predicting car insurance policies using random forest", 2014 10th International Conference on Innovations in Information Technology (IIT), 2014 Publication	
5	Submitted to University of North Texas Student Paper	
6	link.springer.com	

Internet Source

1 %

7 downloads.hindawi.com
Internet Source
<1 %

8 www.hindawi.com
Internet Source
<1 %

9 www.researchgate.net
Internet Source
<1 %

10 Peng Li. "Image Color Recognition and Optimization Based on Deep Learning", Wireless Communications and Mobile Computing, 2022
Publication
<1 %

11 Sourav Ray, Mohaiminul Haque, Kazi Al Rakib, Md. Nazmus Sakib, Md. Masnun Rahman. "Experimental Investigation and SVM-based Prediction of Compressive and Tensile Strength of Ceramic Waste Aggregate Concrete", Journal of King Saud University - Engineering Sciences, 2021
Publication
<1 %

Submitted to University of Teesside
Student Paper

•

First Author Certificate

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

PAPER ID
NECICAIEA2K23096

International Conference on
Artificial Intelligence and Its Emerging Areas
NEC-ICAIEA-2K23
17^{th} & 18^{th} March, 2023
Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Bala Murali Krishna Bodapati , Narasaraopet Engineering College has presented the paper title Earthquake Prediction System using machine learning Algorithm in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17^{th} and 18^{th} March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor	Chief-Convenor	Principal, Patron
Dr.S.V.N.Srinivasu	Dr.S.N.Tirumala Rao	Dr. M. Sreenivasa Kumar

Second Author Certificate

NARASARAOPETA ENGINEERING COLLEGE (AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

PAPER ID
NECICAIEA2K23096

International Conference on
Artificial Intelligence and Its Emerging Areas
NEC-ICAIEA-2K23
17th & 18th March, 2023
Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Mannem Aditya , Narasaraopet Engineering College has presented the paper title Earthquake Prediction System using machine learning Algorithm in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

Third Author certificate

NARASARAOPETA ENGINEERING COLLEGE (AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

PAPER ID
NECICAIEA2K23096

International Conference on
Artificial Intelligence and Its Emerging Areas
NEC-ICAIEA-2K23
17th & 18th March, 2023
Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that R. Vijay Kumar , Narasaraopet Engineering College has presented the paper title Earthquake Prediction System using machine learning Algorithm in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

Fourth Author Certificate

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

PAPER ID
NECICAIEA2K23096

International Conference on
Artificial Intelligence and Its Emerging Areas
NEC-ICAIEA-2K23
17th & 18th March, 2023
Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that Sd.Rizwana, Narasaraopet Engineering College has presented the paper title Earthquake Prediction System using machine learning Algorithm in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor	Chief-Convenor	Principal, Patron
Dr.S.V.N.Srinivasu	Dr.S.N.Tirumala Rao	Dr.M.Sreenivasa Kumar