JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

THE CAVALCADE OF IMMACULATE AND DISPERATE ALGORITHMS FOR DETECTING DISTRACTED EARTHQUAKES EMPLOYING MACHINE LEARNING

[1]Nazeer Shaik, [2]Akshaya.G, [3]Deepthi.T, [4]Manoj.A, [5]Aravind.B

[1]Assistant Professor, [2]UG Student, [3]UG Student ,[4]UG Student ,[5]UG Student
[2,3,4,5] Computer Science and Engineering (CSE), [1]CSE (Data Science)
[1,2,3,4,5]Srinivasa Ramanujan Institute of Technology, Anantapur, India

Abstract: An Earthquake is the sudden shaking of the surface of the earth resulting from sudden release of energy in the Lithosphere. Reliable prediction of earthquakes has numerous societal and engineering benefits. In recent years, the exponentially rising volume of seismic data has led to the development of several automatic earthquake detection algorithms through machine learning approaches. Different algorithms have been applied on earthquake detection like Decision tree, Random Forest classifier, Support vector Machine, K Nearest Neighbour Classifier.

Index Terms - Machine Learning, Earthquake, Random Forest Classifier, Decision tree, Support Vector Machine, K Nearest Neighbour Classifier.

I. INTRODUCTION

Earthquakes generally occur because of plate tectonics. While it is important to understand the nonlinear dynamics of earthquake generation process. Earthquake detection is an essential area of research that deals with identifying and analyzing seismic activity, which can cause significant destruction and loss of life. Traditional earthquake detection methods rely on seismometers, which can be expensive and not always reliable. However, with the advancements in machine learning, earthquake detection using machine learning algorithms has gained popularity. Machine learning algorithms can identify patterns in large amounts of seismic data and detect seismic activity in real-time. This has the potential to improve the speed and accuracy of earthquake detection, allowing for earlier warnings and more effective disaster response. In this context, machine learning algorithms are trained on seismic data and use various techniques to detect earthquakes.

II. LITERATURE SURVEY

Ali G. Hafez, Ahmed Abdel Azim, M. Sami Soliman & Hideki Yayama [1] have proposed a technique called P-wave picking which is one of the important steps for earthquake parameter determination. Although manual inspection of P-wave arrival timing is the most accurate method for detection but, using automated algorithm is a must to facilitate this continuous task. This algorithm generates a daily report of all events recorded by this sub-network. This algorithm can detect very small events starting from microearthquakes due to the use of multiresolution analysis (MRA) of discrete wavelet transform (DWT). Results show a high rate of successful detections of 94.6% with low false alarm rate.

Lomax, A. Michelini and D. Josipovic's [2] have suggested an investigation of rapid earthquake characterization using single station waveforms and a convolutional neural network. Effective early-warning, response and information dissemination for earthquake sand tsunamis require rapid characterization of an earthquake's location, size and other parameters. This characterization is mainly provided by real-time seismogram analysis using established, rule-based, seismological procedures. With the advent of powerful machine learning tools to make predictions from large data sets.

Omar M. Saad, Ali G. Hafez, and M. Sami Soliman [3] has proposed Deep Learning Approach for Earthquake Parameters Classification in Earthquake Early Warning System. Magnitude determination of earthquakes is a mandatory step before an earthquake early warning (EEW) system sends an alarm. Beneficiary users of EEW systems depend on how far they are located from such strong events. Therefore, determining the locations of these shakes is an important issue for the tranquillity of citizens as well. The proposed algorithm depends on a convolutional neural network (CNN) which can extract significant features from waveforms that enabled the classifier to reach a robust performance in the required earthquake parameters. The classification accuracies of the suggested approach for magnitude, origin time, depth, and location are 93.67%, 89.55%, 92.54%, and 89.50%, respectively.

H Hang Zhang 1 2, Jun Zeng 1, Chunchi Ma 1 3, Tianbin Li 1, Yelin Deng 1, Tao Song [4] has proposed a Multi-Classification of Complex Microseismical Waveforms Using Convolutional Neural Network. In this study, a micro seismic multi-classification (MMC)model is proposed based on the short time Fourier transform (STFT) technology and convolutional neural network (CNN). The real and imaginary parts of the coefficients of micro seismic data are inputted to the proposed model to generate three classes of targets. micro seismic data recorded under different geological conditions are also tested to prove the generality of the model, and a micro seismic signal with Mw ≥ 0.2 can be detected with a high accuracy. The proposed method has great potential to be extended to the study of exploration seismology and earthquakes.

Khawaja Asim, Abdul Basit, Francisco Martinez-Alvarez and Talat Iqbal [5] has detected Earthquake magnitude in Hindukush region using machine learning techniques.In the research, four machine learning techniques including pattern recognition neural network, recurrent neural network, random forest, and linear programming boost ensemble classifier are separately applied to model relationships between calculated seismic parameters and future earthquake occurrences. Here, several performance measures can be done with parameters and accuracy can be estimated.

III. PROPOSED WORK

1.Random Forest Classifier

- A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.
- A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.
- The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.
- A random forest eradicates the limitations of a decision tree algorithm. It reduces the over fitting of datasets and increases precision. It generates predictions without requiring many configurations in packages (like Scikit-learn).

Features of a Random Forest Algorithm:

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of over fitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.

2.Decision Tree

Overall, the decision tree algorithm can help improve the accuracy of earthquake detection by identifying the most important features for classification, allowing for the creation of more complex models that capture the nuances of the seismic data. This can ultimately lead to faster and more accurate detection of earthquakes, which is critical for early warning and response systems.

The decision tree algorithm is a method used to identify important features in earthquake detection using Artificial Neural Networks (ANN). It works by creating a tree-like structure that splits the seismic data into branches based on the most informative features, helping to distinguish between earthquake and non-earthquake events. This algorithm can lead to more accurate detection of earthquakes by identifying patterns in the seismic data that may be difficult for humans to recognize. By using both ANN and decision tree algorithms, more complex models can be developed, which can improve the accuracy of earthquake detection and provide faster warnings to people in affected areas.

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome.

3.Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

SVC works by finding the optimal boundary, or hyperplane, between two classes of data points, in this case, earthquake and non-earthquake events. The algorithm identifies the data points closest to the boundary, known as support vectors, and uses them to construct the boundary.

4.K-Nearest Neighbour

K-Nearest Neighbours (KNN) algorithm is a machine learning algorithm that can be used for various tasks, including classification and regression. In the context of earthquake detection, KNN can be used as a predictive model to identify seismic events based on their similarity to previously recorded earthquakes.

In earthquake detection, KNN can be used as a pattern recognition tool to identify seismic events that are similar to known earthquakes based on their location, magnitude, and other features. This can help in early warning systems and rapid response to earthquakes. For example, if a new seismic event occurs, KNN can be used to quickly identify if it is similar to known earthquakes in the area, and if so, issue an alert to nearby populations to take appropriate action.

IV. SYSTEM ARCHITECTURE

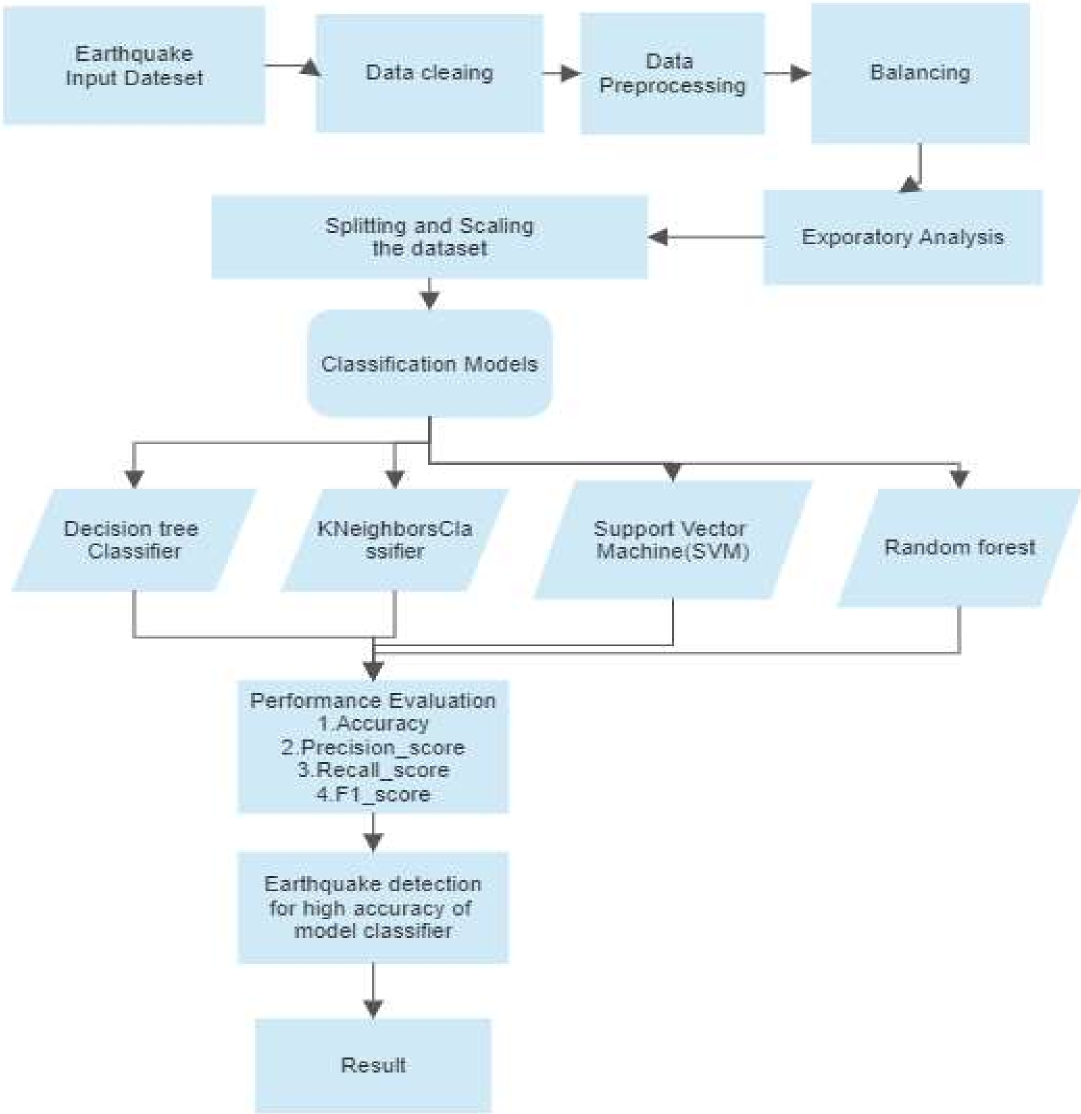

Figure.1: System Architecture

4.1. Steps Involved in Design

- Data Collection
- Data Pre-Processing
- Model Training
- Model Evaluation

4.1.1 Data Collection

Data is an important asset for developing any kind of Machine learning model. Data collection is the process of gathering and measuring information from different kinds of sources. This is an initial step that has to be performed to carry out a Machine learning project. In the present internet world these datasets are available in different websites (Ex: Kaggle, Google public datasets, Data.gov etc.) The dataset used in our project is downloaded from the Kaggle website and it contains nearly 116 records and 20 different attributes. The dataset consists of 19 independent attributes and one dependent attributes. So, the aim of the project is to predict the dependent variables using independent variables.

4.1.2 Data Preprocessing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always the case that we come across clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put it in a formatted way. So, for this, we use data preprocessing tasks. Earthquake Detection using Machine Learning Algorithms.

Preprocessing of the data consists of different kinds of steps in which analysis of the data, Data cleaning, Data encoding are part of this.

4.1.1.1 Exploratory Data Analysis

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always the case that we come across clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put it in a formatted way. So, for this, we use data preprocessing tasks. Preprocessing of the data consists of different kinds of steps in which analysis of the data, Data cleaning, Data encoding are part of this.

- Dimension reduction techniques which help create graphical display of high dimensional data containing many variables.
- Univariate visualization of each field in the raw dataset, with summary statistics.
- Bivariate visualizations and summary statistics that allows you to assess the relationship between each variable in the dataset and the target variable in the dataset and the target variable you're looking at.
- Multivariate visualizations, for mapping and understanding interactions between different fields in the data.
- This data analysis is of two types:
  a. Univariate analysis
  b. Bivariate analysis
    - Univariate analysis is the simplest form of data analysis where the data being analyzed contains only one variable. Since it's a single variable it doesn't deal with causes or relationships.
    - Bivariate data is data that involves two different variables whose values can change. Bivariate data deals with relationships between these two variables.

4.1.1.2 Filling Missing Data & Data Encoding

- The next step of data preprocessing is to handle missing data in the datasets. If our dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset.
- By calculating the mean and Mode: In this way, we will calculate the mean or Mode of that column or row which contains any missing value and will put it on the place of missing value. This strategy is useful for the features which have numeric data such as age, salary, year, etc.
- Data encoding: Since the machine learning model completely works on mathematics and numbers, but if our dataset would have a categorical variable, then it may create trouble while building the model. So, it is necessary to encode these categorical variables into numbers.
- Our dataset also consists of different categorical data in which they are encoded in this step.

4.1.3 Training the Model

In this step the model is trained using the algorithms that are suitable. Flood prediction is a kind of problem in which One variable has to be determined using some independent variables Regression model is suitable for this kind of scenario.

- A training model is a dataset that is used to train an ML algorithm. It consists of the sample output data and the corresponding sets of input data that have an influence on the output. The training model is used to run the input data through the algorithm to correlate the processed output against the sample output.
- Our project implements these algorithms like Logistic Regression, K Nearest Neighbor, Support Vector Machine, Random Forest.

4.1.4 Model Evaluation

In this step the trained model is evaluated by determining the accuracy of the model against the test data. various ways to check the performance of our machine learning or deep learning model and why to use one in place of the other. We will discuss terms like:

- Accuracy
- Recall score
- Precision score
- F1 score

Out of these we used Accuracy for evaluating our model. Accuracy is the most commonly used metric to judge a model and is actually not a clear indicator of the performance. The worst happens when classes are imbalanced balanced datasets. In such cases, other evaluation metrics such as precision, recall, and F1 score can provide a more informative picture of the model's performance. It is important to carefully choose the evaluation metrics based on the characteristics of the dataset and the problem being solved. In the case of imbalanced datasets, using accuracy alone can lead to inaccurate conclusions about the performance of a machine learning model, and it is important to consider alternative metrics such as precision, recall, and F1 score.

V. RESULTS AND DISCUSSION

5.1 Evaluation of Algorithms:

Evaluates the performance of four different machine learning algorithms, including decision tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest.

	Name	Accuracy	Precision_score	Recall_score	F1_score
0	DecisionTreeClassifier	0.942308	0.941710	0.942308	0.941875
1	KNeighborsClassifier	0.923077	0.934314	0.923077	0.923851
2	SVM	0.930769	0.945339	0.930769	0.932321
3	RandomForestClassifier	0.976923	0.977345	0.976923	0.976927

Table 5.1: Evaluation of Algorithms

5.2 Performance comparison of Model:

After evaluating the performance of the four machine learning algorithms, the results can be compared to determine which algorithm performs the best. For instance, the accuracy, precision, recall, and F1 score.

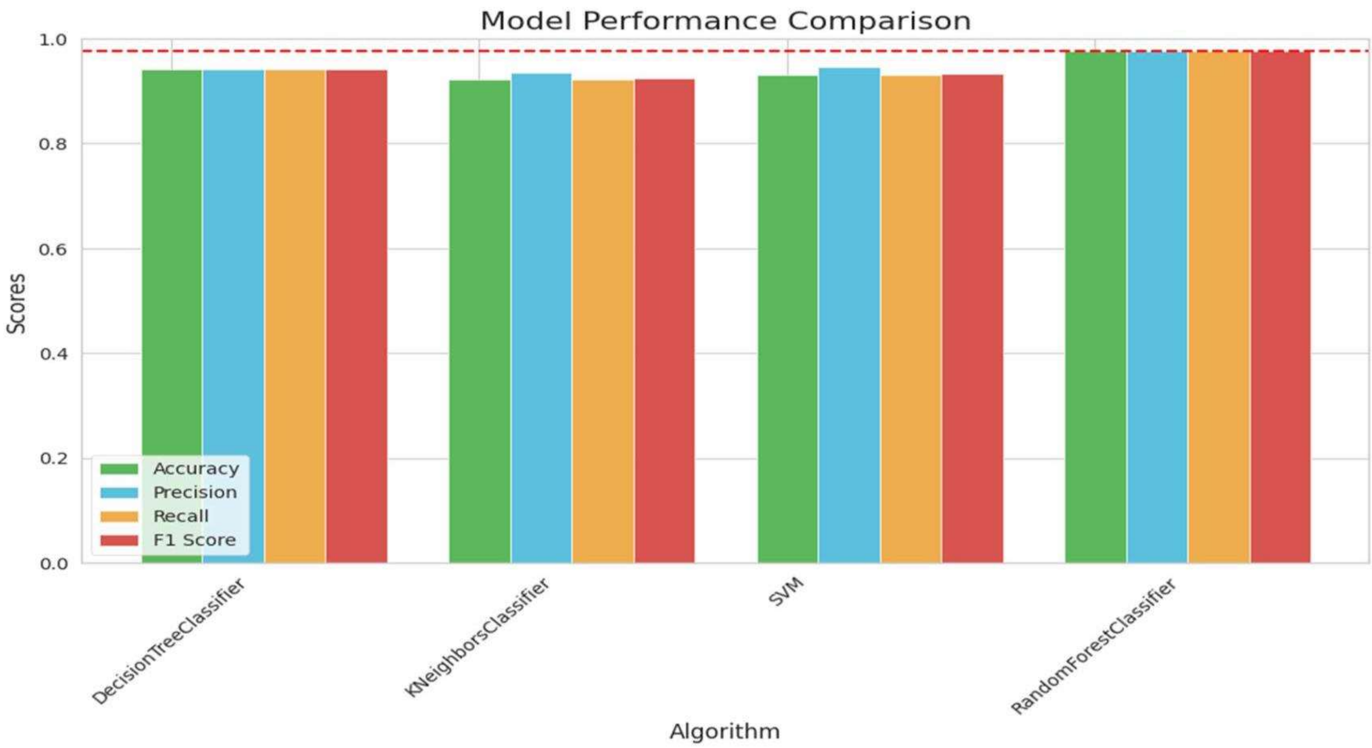

Table 5.2: Performance Comparison of Algorithms

5.3 Web app results:

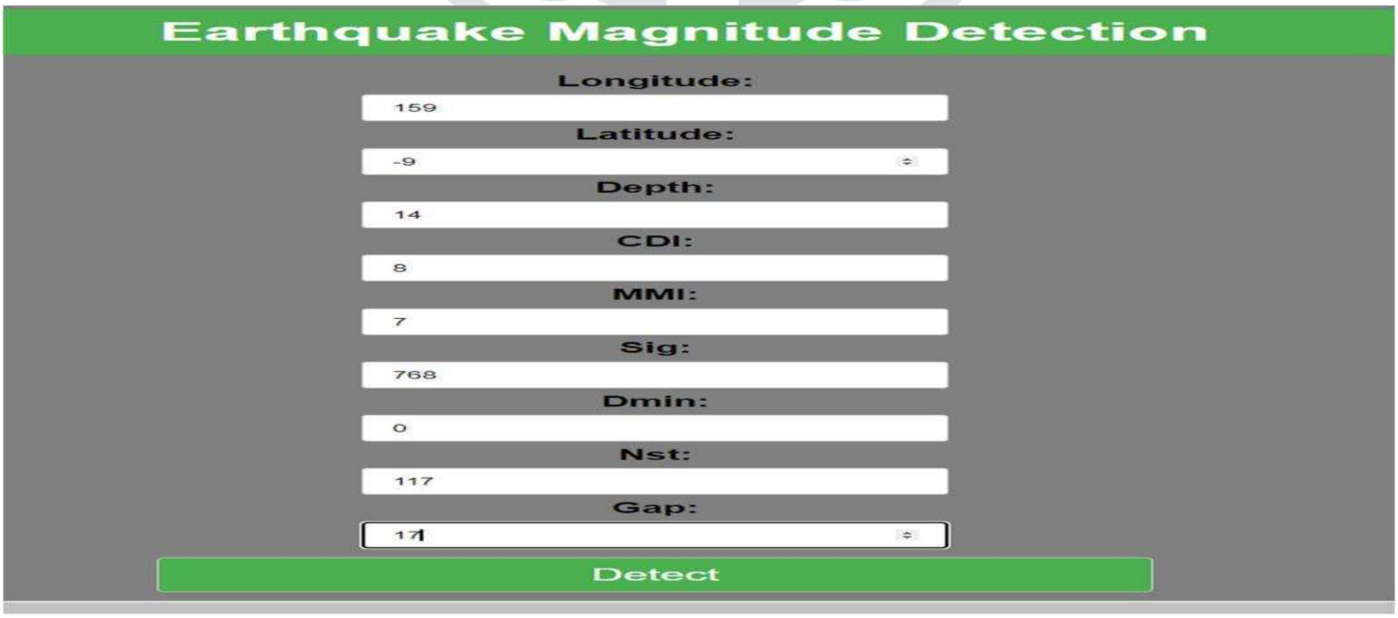

Fig 5.3: Detection of earthquake using web app UI

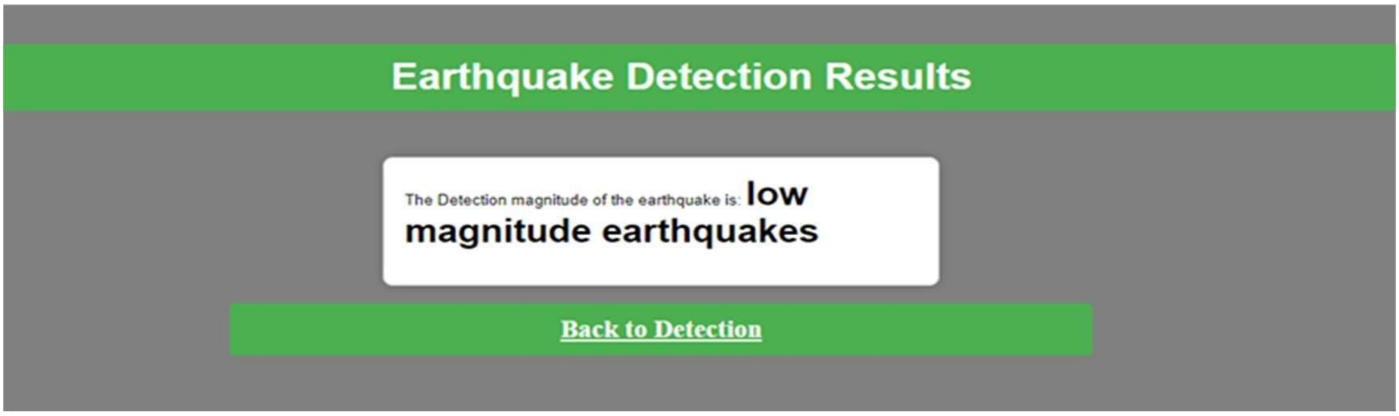

Fig 5.4: Result of the earthquake using webapp UI

V. CONCLUSION

Based on the evaluation of four machine learning algorithms, namely Decision Tree Classifier, Kneighbors Classifier, SVC, and Random Forest Classifier, it can be concluded that all models have performed well in detecting earthquakes. Among all the algorithms, the Random Forest Classifier algorithm achieved the highest accuracy score of 0.976923, which is the best performance compared to the other models. The Decision Tree Classifier, Kneighbors Classifier, and SVC algorithms also performed well, with accuracy scores of 0.942308, 0.923077, and 0.930769, respectively. Moreover, all the models have performed well in terms of precision, recall, and F1-score. These performance metrics are essential for earthquake detection as it requires accurate and reliable predictions to avoid false positives and false negatives. In conclusion, the Random Forest Classifier algorithm is the best model for earthquake detection based on the given dataset. However, other algorithms such as Decision Tree Classifier, K-Neighbour's Classifier, and SVC can also be used as they have also shown good performance.

VI. REFERENCES

1. Dinkytulsi Nandwani, 2vanita Buradka, Earthquake Damage Prediction Using Machine Learning, © 2022 Ijcrt | Volume 10, Issue 7 July 2022 | Issn: 2320-288, Www.Ijcrt.Org.

2. Roxane Mallouhy, Chady Abou Jaoude ,Christophe Guyeu, Abdallah Makhoul, Major earthquake event prediction using various machine learning and author profiles for this publication at: https://www.researchgate.net/publication/339901560.

3. Vindhya Mudgal, 2jayashree M Kudari, 3ravi Chandra A,Prediction Of Earthquakes Using Machine Learning Algorithms: A Survey Paper" Issn :2349-5162 ,Estd Year 2014,Jetir.Org.

4. Dr. S. Anbu Kumar1, Abhay Kumar2, Aditya Dhanraj3, Ashish Thakur, Earthquake Prediction using Machine Learning, ISO 9001:2008 Certified Journal, International Research Journal of Engineering and Technology (IRJET), Issue: 05 | May 2021.

5. Alyona Galkina, Natalia Grafeeva, Machine Learning Methods for Earthquake Prediction: a Survey,CEUR –ws.org/vpl-2372/SEIM_2019_paper_31.

6. W. Li, N. Narvekar, N. Nakshatra, N. Raut, B. Sirkeci and J. Gao. "Seismic Data Classification Using Machine Learning", 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService), Bamberg, 2018, pp. 56-63.

7. G. Cortés, A. Morales-Esteban, X. Shang, and F. Martínez-Álvarez, "Earthquake Prediction in California Using Regression Algorithms and Cloud-based Big Data Infrastructure," Computers & Geosciences, vol. 115, pp. 198-210, 2018.

8. Pratiksha Bangar, Deeksha Gupta, Sonali Gaikwad, Bhagyashree Marekar, Jyoti Patil, "Earthquake Prediction using Machine Learning Algorithm" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-8 Issue-6,March 2020.