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

PREDICTION OF EARTHQUAKES USING MACHINE LEARNING ALGORITHMS: A SURVEY PAPER

[1]Vindhya Mudgal, [2]Jayashree M Kudari, [3]Ravi Chandra A

[1]UG Scholar, [2]Associate Professor, [3]UG Scholar
[1,2,3]Bachelor Of Computer Application,
[1,2,3]School of CS and IT, Jain (Deemed to be University), Bangalore, India

Abstract : Earthquakes are by far the most dangerous natural disasters because they can strike without warning, providing little time to adapt. Because of this, the issue of earthquake prediction is crucially essential for human safety. Despite the scientific community's continuous interest in this subject, no agreement exists on whether or not viable to produce a solution that includes enough accuracy. However, the machine learning's successful application, methods to other areas of study shows that they may be utilized to be able to create more precise short-term forecasts. This paper explores recent studies that studied the use of numerous machine learning-based algorithms to earthquake prediction. The purpose is to organize the approaches deployed and research the significant developments in forecasting. We are convinced that our research will be informative and encouraging to both Scientists studying earthquakes as well as novices to the area [1].

***IndexTerms* - MapReduce, SVM, Adaboost, Decision Tree, Random Forest, Forecasting.**

I. INTRODUCTION

An earthquake is described as ground movement and shaking that causes energy to be released from rocks. Earthquakes, like many other natural catastrophes, produce widespread destruction, financial loss, and human harm. Earthquakes occur on a daily basis all around the planet. Many processes and occurrences impacting many aspects of life have been adequately explored to offer predictions. Risk analysis allows us to evaluate if an event is likely to occur at a certain moment in time, as well as respond to it immediately or even prevent it from occurring. Even in the contemporary world, however, people have an effect on some occurrences. Tsunamis, tornadoes, floods, volcanic eruptions, and other natural catastrophes are examples of natural disasters. An oncoming threat cannot be stopped by humans; but, precautions and timely response are available to lessen economic and human losses. Not all natural disasters, however, have been carefully explored and are "predictable." For starters, they typically strike without prior notice, giving individuals inadequate time to take preparations. Earthquakes frequently create tsunamis, snow falls, and landslides. They have the capability of causing industrial disasters.

There is presently no conventional technique for earthquake prediction. Furthermore, scientists are divided on whether a solution to this challenge is even attainable. Machine learning methods' fast progress, as well as their successful application to a wide range of situations, implies that these technologies might be effective in discovering hidden patterns and producing accurate predictions [1].

II. OBJECTIVES

Our purpose is to analyze the available methodologies on prediction of earthquakes using machine learning algorithms which enables us to know the algorithm that gives the greatest accuracy by comparing the algorithms with one another.

III. BACKGROUND

This section examines a number of publications that investigated the use of machine learning methods to the task of earthquake prediction over various temporal and spatial intervals [1].

The MapReduce model, K-Nearest Neighbours, Logistic regression, Naive Bayes, Support Vector Machine, Decision tree algorithm-based method, shallow machine learning, deep learning algorithms, Random Forest, PNN model, Adaboost, Classification and Regression trees are all reviewed.

3.1 MapReduce Model - Currently, big data analytics is among the fastest growing industries. The Hadoop framework is made up of two parts: For parallel processing, map Reduce computation and Storage is handled via HDFS. Hadoop has its own library for distributed file systems, to which they can gain access, with name nodes and data nodes as part of its conceptual representation. The Map function distributes conventional data analysis by distributing given tasks to various nodes in the Hadoop ecosystem. In general, the Map Reduce model addresses a large number of issues at the same time [7].

Data analysis using Apache Hadoop is a simple and efficient method. Hadoop is constructed in such a way to operate in a distributed setting. handles a group of diverse nodes that perform the reduce and map functions. This system performs data analysis on earthquake data by year and location. The data is utilized to determine the most shaking area, the location closest to the line that is incorrect and the shakings per minute at the moment [7].

3.2 KNN - The K-nearest neighbours (KNN) approach can be used for both classification and regression problems, however it is most commonly employed to forecast classification challenges. KNN is founded on the idea that cases with similar features occur frequently. By looking at the class of its closest neighbours, the unclassified instance's class value can be expected. The KNN technique locates the query instance's K nearest neighbours, which is the only variable in this algorithm, then classifies it by selecting the most common class label. K can be changed to make the model more or less flexible [4].

3.3 SVM (Support Vector Machine) - Regression is a well-known deep learning approach that relies on kernel algorithms to provide accurate predictions. SVM is a machine learning technique that was created to prevent ANN deficits. SVM provides optimum. The SVM approach, by employing nonlinear functions, the sample space is transformed into a high-dimensional Eigenspace. Because seismic occurrences are nonlinear, SVM aids in precise prediction [9].

The SVM method has been modified to become LS-SVM. Although The SVM technique is simplified using LS-SVM requires kernel parameters, which are critical for regression difficulties. As a result, we need a suitable method for the optimum selection of LS-SVM parameters, which is accomplished by optimizing LS-SVM with FPA [3].

3.4 Logistic Regression - The logit model, often known as logistic regression, evaluates the connection between one dependent variable and several independent factors. It uses a logit function to forecast the likelihood of an occurrence. (1) binary: there are just two conceivable outcomes, for example, good or terrible. (2) Multinomial: a collection of three or more unconnected categories [4].

3.5 Naive Bayes Algorithm - According to Thomas Bayes' work, the Naive Bayes method is a drastically oversimplified model that is probabilistic quantifies counting probabilities permutations in a data set of values and frequency collection. The Classification based on Bayesian theory method suppose the data belongs to a specific class, then calculates the likelihood that the hypothesis is correct. The Bayesian Naive model is built on the strong assumptions of freedom, which states that changing the conditional distribution of one attribute does not impact the likelihood of the other [4].

3.6 Decision Tree - The most popular method is to use a decision tree, powerful and widely used classification and prediction approach. The fundamental drawbacks of decision trees are that their performance is too reliant on the initial sample, and they have a slow learning rate. Several trees are utilized to solve these challenges [4].

3.7 Shallow Machine Learning - There are three types of shallow ML approaches: classical ML, NN-based techniques and clustering approaches For earthquake prediction, traditional machine learning SVM, SVR, KNN, RF, and DT are examples of algorithms, and others Feature handcrafted elements. To keep similar studies together, there will be two categories in this section. The two kinds of earthquake and aftershock prediction research for shallow machine learning-based approaches are earthquake characteristics studies and earthquake prediction studies for shallow machine learning-based techniques [9].

3.8 Deep learning algorithms - This machine learning procedure does not require handmade characteristics and the ability to produce thousands of highly features of distinction that are tough to find by hand. Multiple hidden layers are used in these DL-based models, which can be time-consuming. These models may be prone to overfitting due to their advanced features. As a result, the concepts of dropout and regularization are employed. The features that a fully connected layer is utilised in classification are found using multiple hidden layers. We'll divide the research into two categories: earthquake characteristics research earthquake and aftershock prediction research for deep learning-based approaches; and deep learning-based approaches for earthquake and aftershock prediction [9].

3.9 Random forest - Several trees are utilized to solve performance issues that are unduly based on the original sample and a star topology that might vary dramatically with the addition of the new survey. To prevent Randomness is created by having equal trees introduced: Each tree has a different view of the situation generated randomly from the source conclusions and observations of explanatory factors. The tree bagged algorithm is the process of constructing tree structure selected according to a random selection from among the data [4].

3.10 PNN model –The Parzen windows classifier principle and its application to Bayesian statistics are the foundations of the PNN. PNN architecture for forecasting the magnitude of the biggest event in the next time period consists of an input layer, a hidden layer called the pattern layer, a second hidden layer called the summation layer, and a single-node output layer called the competition layer. Each of the eight nodes in the input layer represents a seismicity indication. The pattern layer has the same number of nodes as n represents the number of training input vectors that are available. N nodes make up the summation layer, which is the same as the number of input classes [2].

3.11 Adaboost - A meta-algorithm employs a series each has access to a set of weak algorithms each has a set of ineffective algorithms at its disposal issues, focusing on tough observations and driving its successors to handle them appropriately. In a wide sense, the phrase boosting refers to strategies based on the notion of serial assembling of weak classifiers [4].

3.12 Classification and Regression Trees (Cart) - A decision tree is a tool for explaining a value based on the set of continuous or discrete factors. These are reasonably effective, non-linear and non-parametric techniques of dividing individuals into groups [4].

3.13 Hybrid-Machine Learning:

3.13.1 ELM - ELM (Extreme Learning Machine) is a simple method that relies on a feed-forward neural network. With just one hidden layer, the data travels in one direction (forward). In the classification and regression sectors, ELM is a well-known and well-proven technique. A sigmoid function is often used in ELM [3].

This job is separated into four parts, the first of which is the processing of raw data to detect seismic indicators. The second stage is to use machine learning algorithms to process these signals, which will anticipate or output the predicted magnitude in the future. The final step is to use artificial intelligence tools to optimize the output. The fourth section involves comparing different outcomes obtained via the use of various methods [3].

3.13.2 The LS-SVM - It is a deep and supervised learning algorithm capable of classifying input and predicting values. SVM has a new version called LS-SVM, that addresses SVM problems. Instead of the Quadratic programming issue utilized in SVM, the solution in LS-SVM may be established by solving several linear equations [3].

3.13.3 FPA - FPA also called the Flower Pollination Algorithm is an optimization approach based on the concept of flower pollination. Pollination occurs in two ways: biotic pollination, which occurs in 90% of flowers due to pollen dispersal by various insects, and abiotic pollination, which occurs in 10% of flowers due to pollen movement through water or wind [3].

FPA optimizes the ELM algorithm to provide us with an ideal forecast of the occurrence of the Earthquake, as well as improving LS-SVM with FPA to improve the accuracy of earthquake magnitude prediction [3].

IV. RESULTS

There is currently no generic approach for forecasting earthquakes with acceptable accuracy [1]. Machine learning methods have improved prediction accuracy more [6]. There is no ideal model that produces 100% accurate results, m global solutions while avoiding local minima issues [3].

It creates a hyperplane on an N-dimensional plane to arrange the classes that ensures the maximum feasible margin distance between the data points of the classes. A higher-dimensional nonlinear hyperplane is required when a linear hyperplane fails to partition the classes, but at least a trial is conducted to increase the accuracy as much as feasible [3]. The authors trained the ANN in two rounds (first on outliers, then on the whole training dataset) to increase its performance on large quakes, and the resultant accuracy rate was 58.02 percent [1]. The PNN model developed in this study meets excellent prediction accuracies for earthquakes with magnitudes ranging from 4.O to 5.5, but does not reach satisfactory findings for earthquakes with magnitudes greater than 5.5 [2]. The experimental findings indicated that after using the FPA method to optimize it, ELM and LS-SVM accuracy both improved. According to all criterion comparisons, the suggested In comparison to the FPA-ELM model, the FPA-LS-SVM model fared better. In addition, In earthquake prediction, FPA-LS-SVM is the most effective in lowering the false alarm ratio[3].Random Forest performs the best in terms of accuracy, with 76.97 percent, closely followed by KNN had 75.53 percent, MLP had 74.82 percent, and SVM had 74.82 percent. SVM, MLP, and AdaBoost all have a 72.66 to 74.82 percent accuracy, whereas CART predicts 70.5 percent. However, Naive Bayes and Logistic Regression had the lowest prediction percentages of 66.9 percent and 66.9 percent, respectively [4].

The random forest-support vector machine approach works well together for huge datasets. The accuracy of this stacking model is the highest—83 percent [5]. The suggested system was built utilizing the Apache Hadoop framework. Data analysis with Apache Hadoop is a quick and easy process. Apache Hadoop is designed to work in a distributed setting, and it manages a collection of different nodes that perform map and reduce operations. The result suggests that the next likely location of earthquake was appropriately recognized. Furthermore, the findings are utilized to determine the shakiest area, the position closest to the defective line, and the present location shakes per minute [7]. The accuracy of the KNN algorithm is 76.92 percent, and the accuracy of the logistic regression algorithm is 61.54 percent. The Decision tree algorithm has an accuracy rating of 76.92 percent [8]. The optimal set of seismic parameters is not determined for earthquake prediction investigations. Different parameter sets are used in different research. The amount of completeness varies according to the dataset. The 99.49% accurate BP-AdaBoost model, was the best [9].

TABLE I. SUMMARY OF RESULTS.

Algorithm name	Accuracy		Ref.
ANN	58.02		[1]
RANDOM FOREST	76.97		[4]
KNN	75.53	76.92	[4],[8]
MLP	74.82		[4]
SVM	74.82		[4]
CART	70.5		[4]
NAÏVE BAYES	66.9		[4]
LOGISTIC REGRESSION	66.9	61.54	[4],[8]
RANDOM FOREST -SVM	83		[5]
DECISION TREE	76.92		[8]
BP-ADABOOST	99.49		[9]

V. CONCLUSION

This paper discusses the authors' earthquake prediction experiments. AI-based approaches have opened up new opportunities for enhancing the prediction process due to their greater precision when compared to conventional procedures. The BP-AdaBoost model, which was 99.49 percent accurate, came out on top, followed by random forest and the support vector machine stacking model, which was 83 percent accurate. Each machine learning approach yields outcomes that differ from one another.

VI. ACKNOWLEDGMENT

The successful completion of any task would be incomplete without thanking the people who made it possible whose constant guidance and encouragement motivated my effort with success. It is our great pleasure to thank our guide, Prof Jayashree M Kudari for her excellent guidance, constant encouragement, support, constructive suggestions. We also thank all the faculties of the Computer Application Department for their suggestions that enabled us to surpass many of the seemingly impossible hurdles. We would also like to thank all the staff members of the Department of Bachelor of Computer Applications-Data Analytics for their inspiration and kind cooperation in completing the paper. We are also grateful to all the university staff for their consistent support and assistance.

REFERENCES

[1] Alyona Galkina, et al, “Machine Learning Methods for Earthquake, Prediction: a Survey” Saint Petersburg State University, Saint Petersburg.

[2] HUANG Sheng-Zhong, “The prediction of the earthquake based on neutral networks.”, International Conference On Computer Design and Applications (ICCDA 2010),2010.

[3] Mustafa Abdul Salam,Lobna Ibrahim, Diaa Salama Abdelminaam ," Earthquake Prediction using Hybrid Machine ,Learning Techniques" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 5, 2021.

[4] Roxane Mallouhy, Chady Abou Jaoude, Christophe Guyeux, Abdallah Makhoul, "Major earthquake event prediction using various machine learning algorithms"

[5] 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, Volume-8 Issue-6, March 2020.

[6] Rabia Tehseen, Muhammad Shoaib Farooq and Adnan Abid, "Earthquake Prediction Using Expert Systems: A Systematic Mapping study”,12, 2420, Sustainability 2020.

[7] C.P.Shabariram and Dr. K.E.Kannammal,"EARTHQUAKE PREDICTION USING MAP REDUCE FRAMEWORK", 2017 International Conference on Computer Communication and Informatics (ICCCI -2017), Coimbatore, INDIA ,2017.

[8] Wanjiang Han, Yuanlin Gan and Shuwen Chen and Xiaoxiang Wang, "Study on Earthquake Prediction Model Based on Traffic Disaster Data" , 978-1-7281-6579-0/20/$31.00©2020 IEEE,2020

[9] Md. Hasan Al Banna, Kazi Abu Taher,M. Shamim Kaiser ,Mufti Mahmud , Md. Sazzadur Rahman,A. S. M. Sanwar Hosen And Gi Hwan Cho ,"Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges" , IEEE Access, Volume 8,2020.