**EARTH DAMAGE PREDICTION MODEL**

**US**

1V. N. V. L. S. Swathi

Department of CSE

Name of organization: CVR College of Engineering

City: Hyderabad, India.

E-mail: swathivllr@gmail.com

2 K. Kalyan, 3 G. Karthik SaiTeja, 4 C. Keerthan

Department of CSE

Name of organization: CVR College of Engineering

City: Hyderabad, India.

***Abstract***

An earthquake is a natural disaster known on account of the devastating effect it has on naturally occurring structures and manmade structures such as buildings, bungalows and residential locations to name a few. Earthquakes are measured using seismometers, that detect the vibrations due to seismic waves travelling through the earth’s crust. In this work, the damage that is caused by an earthquake was classified into damage grades, ranging in values from one to five. A previously acquired data set was used, wherein a series of parameters were taken into consideration to predict the damage grade of a given building, which is associated with a Unique Identification String. The prediction was done using a survey of existing machine learning classifier algorithms

**Keywords**: earthquake,damage prediction,building damage,machine learning.

1. INTRODUCTION

An earthquake is a calamitous occurrence that is detrimental to human interest and has an undesirable impact on the environment. Earthquakes have always caused in calculable damage to structures and properties and caused the deaths of millions of people throughout the world. In order to minimize the impact of such an event, several national, international and transnational organizations take various disaster detection and prevention measures. Time and quantity of the organization’s resources are limiting factors, and organization managers face several difficulties when it comes to the distribution of the resources.Leveraging the power of machine learning is a viable option to predict the degree of damage that is done to buildings.

It can help identify safe and unsafe buildings which helps to predict damage prone areas and thus avoiding death and injuries resulting from an earthquake, while simultaneously making rescue efforts efficient.This is done by classifying these structures on a damage grade scale based on various factors like its age, foundation, number of floors, material used and several other parameters. Then the number of families and the probable casualties ward-by-ward in a district are taken into account.This enables distribution of relief forces proportionately ward-wise and its prioritization based on the extent ofdamage.

Earthquake are quite fatal and can cause quite a loss. Those that occur in the workplace can cause harm to employees, environment and damage to the equipment. Industrial related accidents, injuries and fatality data demonstrate that continued efforts and effective measures are necessary to reduce the number of industrial accidents, illnesses and and fatalities.This prediction can help identify safe and unsafe buildings which helps to predict damage prone areas and thus avoiding death and injuries resulting from an earthquake, while simultaneously making rescue efforts efficient.

1. LITERATURE REVIEW

In [1], a quick assessment method for earthquake emergency is introduced. The method contains two different modes to obtain damage information from remote sensing images, one of which is based on damage index and the other adopts image classification. The damage index mode relies on traditional visual interpretation. After the damage index is given by experts, the ground intensity data can be gained, and then loss estimate parameters will be acquired from the experiential vulnerability matrix. The image classification mode is an application of digital image processing technique. Those loss estimate parameters can be calculated from the classification result which is sorted by the type of buildings and ranged by the damage degree. While the assessment models are introduced, the action of multi-resourced estimate data is explained to show how to find parameters in various data.

In [2], a probabilistic aspect of the earthquake prediction was given. A theoretical analysis of the earthquake prediction problem in space–time is presented. We find an explicit structure of the optimal strategy and its relation to the generalized error diagram. This study is an extension of the theoretical results for time prediction. The possibility and simplicity of this extension is due to the choice of the class of goal functions. The generalized error diagram allows us to suggest a natural measure of prediction efficiency at the research stage.

In [3], core idea of this work is to predict whereas an event is classified as negative or positive major earthquake by applying different machine learning algorithms. It is well known that there is no best algorithm or one solution that fits all the problems and datasets for machine learning since the performance of algorithms depends on many factors. Some algorithms are best for small data, while others perform better for a tons of data sample. Some algorithms require categorical inputs, while others need quantitative. Another important criterion while choosing the algorithm is the complexity of the dataset and how many features the model needs to learn and predict. This is why, in this work, eight different algorithms have been applied on an earthquake dataset, namely: Random Forest, Naïve Bayes, Logistic Regression, MultiLayer Perceptron, AdaBoost, K-nearest neighbors, Support Vector Machine, and Classification and Regression Trees. For each selected model, various hyperparameters have been tested, and obtained prediction results have been fairly compared using various metrics, leading to a reliable prediction of major events for 3 of them.

1. MACHINE LEARNING TECHNIQUES

Three machine learning techniques have been used in this paper.

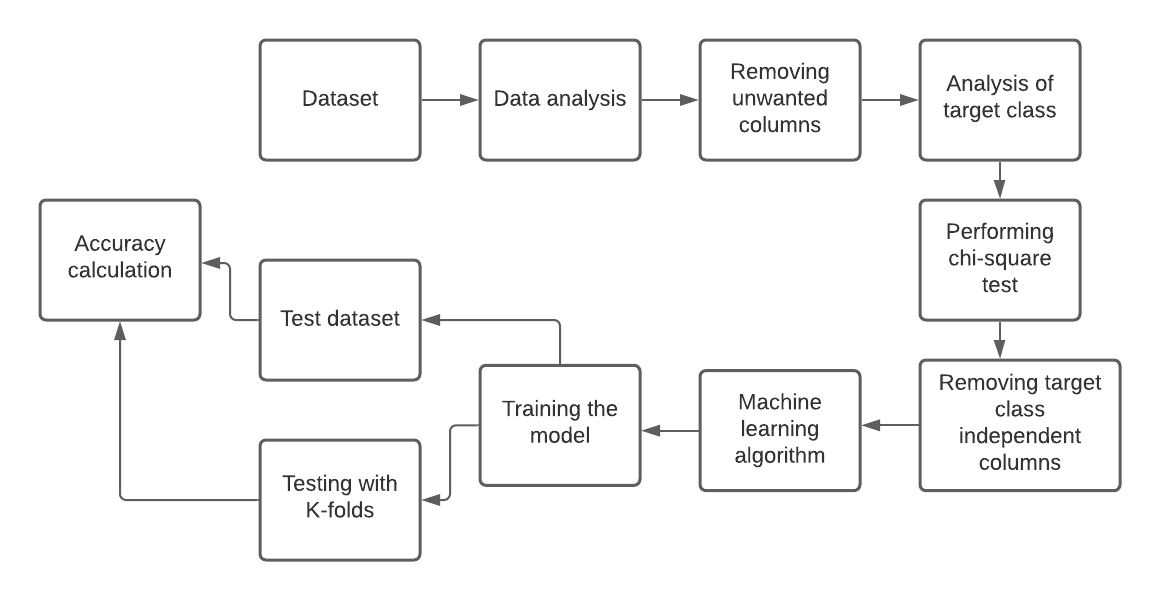
Decision Tree: Decision tree is a supervised learning algorithm which is mostly used for classification problems. Decision tree is used for both categorical and continuous dependent variables. In this algorithm, we divide the population into two or more similar sets. A decision tree is a graphical representation that makes use of branching methodology to represent all possible outcomes of a decision based on certain conditions

Naïve Bayes: Naive Bayes method is a classification technique based on Bayes’ theorem. It is an assumption of independence between predictors. In simple term, a Naive Bayes classifier predicts that the presence of a particular feature in a class is unrelated to the presence of any other feature. [4]

Random Forest:Random forest is a machine learning algorithm that uses a collection of decision trees providing more flexibility, accuracy, and ease of access in the output. This algorithm dominates over decision trees algorithm as decision trees provide poor accuracy as compared to the random forest algorithm. In simple words, the random forest approach increases the performance of decision trees. It is one of the best algorithm as it can use both classification and regression techniques. Being a supervised learning algorithm, random forest uses the bagging method in decision trees and as a result, increases the accuracy of the learning model.

1. IMPLEMENTATION

The below architecture diagram describes the proposed system for EarthQuake damage prediction.



**Figure -1 Architectue diagram**

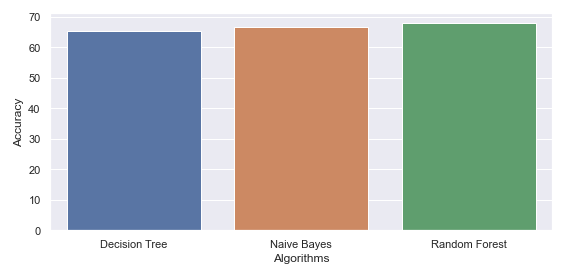
Implementation is in 7 phases

1. Importing required Modules.
2. Extracting the data from Kaggle.
3. Analyzing the data using Exploratory Data Analysis.
4. Removal of unwanted columns .
5. Splitting train and test data.
6. Implementation of different models.
7. Visulazing confusion matrix.
8. METHODOLOGY

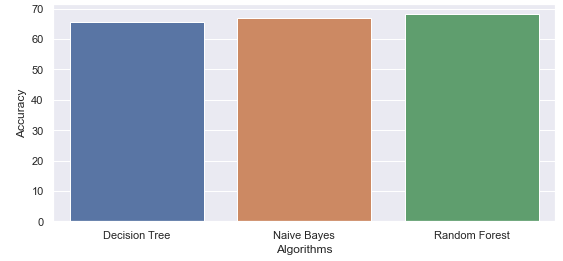
In this paper, we compared the efficiency of different Machine Learning classification algorithms on a particular dataset. The prediction is made on the basis of the physical parameters of the building structure. In validation, we check whether the predictions are accurate or not. We can plot the results and compare them with the actual values, i.e., calculate the distance between the predictions and actual values. Lesser this distance, more accurate will be the predictions. Python is a popular platform used for scientific research and development of production systems. It is a language with number of modules, packages and libraries. Python and its libraries like NumPy, SciPy, Scikit-Learn, Matplotlib are used in data science and data analysis. They are also broadly used for creating scalable machine learning algorithms. Python implements machine learning techniques such as Classification, Regression, Recommendation, and Clustering [13]. After data preprocessing, a dataset is created by splitting it into two parts: 80% of which is used to train the models and 20% is held back as a test dataset. Both training dataset and test dataset are One-hot encoded, i.e., nominal variables are converted into numerical form to be provided to different Machine Learning algorithms for effective prediction. K is Chosen for K-fold cross validation to estimate the accuracy of different models.

1. RESULTS

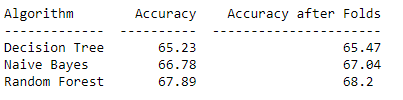
The experiment was run on a system with Microsoft Windows 10 operating system with the configuration of 64GB RAM and 4 Intel cores. Tool python3 was used to run different Machine Learning algorithms. matplotlib library was used to visualize the inner working of the model. We used the accuracy measure for evaluating the quality of the classifier. The purpose of accuracy is to achieve a higher value. We used three different machine learning algorithm, Decision Tree, Naive Bayes, Random Forest on the same dataset. These classification algorithms were run in python with and without a cross validation of 10 folds. The final classification accuracy is considered and compared with each other.



**Figure -2 Accuracy without folds**



**Figure -3 Accuracy with 10 folds**



**Table-1 Accuracy Table**

It is seen Random forest has more accuracy compared to all other.

1. CONCLUSION

Thus, the aim of this project is to predict the damage grade of buildings. The Random forest algorithm is high accurate in predicting compared to Naive bayes and Decision Tree.The use K-folds the accuracy of prediction is increasing the accuracy is directly proportion to number of folds.It is also seen that damage grade 1,5 have high accuracy in prediction.

1. REFERENCES

[1]Long Wang, Xiaoqing Wang, Aixia Dou, Dongliang Wang “Study on construction seismic damage loss assessment using RS and GIS” International Symposium on Electromagnetic compatibility, 2014.

[2]Roxane Mallouhy, Chady Abou Jaoude,Christophe Guyeux ,Abdallah Makhoul,“Earthquake event prediction using various machine learning algorithms”.

[3]G. Molchan and V. Keilis-Borok,“Earthquake prediction: probabilistic aspect”.

[4]<https://www.dezyre.com/article/common-machine-learning-algorithms-for-beginners/202>

[5[]https://www.w3schools.com/python/python\_intro.asp](%5dhttps:/www.w3schools.com/python/python_intro.asp)

[6] H Takata, H. Nakamura, T Hachino,“On prediction of electric power damage by typhoons in each district in Kagoshima Prefecture via LRM and NN”.

[7]Ramli Adnan. Abd Manan Samad, Zainazlan Md Zain, Fazlina Ahmat Ruslan,“5 hours flood prediction modeling using improved NNARX structure: case study Kuala Lumpur”

[8]Dezhang Sun, Baitao Sun,“Rapid prediction of earthquake damage to buildings based on fuzzy analysis”2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, vol. 3, p. 1332-1335, 2010.