IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

EARTHQUAKE DAMAGE PREDICTION USING MACHINE LEARNING

[1]Dinky Tulsi Nandwani, [2]Vanita Buradkar

[1]Student, [2]Professor

[1]Computer Science & Engineering,

[1]Rajiv Gandhi College of Engineering & Technology, Chandrapur, India

Abstract: Earthquake is one of the most risky herbal calamities. It happens because of unexpected shaking of floor that is because of motion of seismic waves. When the strain on the brink overcomes the friction, there is an earthquake that releases energy in waves that travel via the earth's crust and motive the shaking that we experience. The project Predicting damage to a building during earthquake is made so that classification of building grades can be done. The building grade is divided into five categories: minor, medium, major, partial collapse & total collapse. If the grade of the building is low, then it will cause more damage to building during earthquake. If the grade of the building is high, then it will cause less damage to a building during earthquake. Therefore by knowing grade of the buildings, necessary precautions may be taken before earthquake**.**

Index Terms - Earthquake, Calamities, Seismic Waves, Building Grade, Predicting.

I. INTRODUCTION

The Project "Predict the damage to a building" is made for reducing the destruction and damage caused to buildings by earthquakes. Earthquakes are as a result of the actions of tectonic plates of the Earth. When the two plates collide at the fault line, earthquake occurs and it causes the violent shaking of the ground and buildings. [1] Earthquakes occur suddenly and at any time in day or night. Small earthquakes does not cause much damage but large earthquakes cause lot of damage and destruction to homes, roads, bridges, buildings etc. and also causes deaths of human beings.

Many measures are taken to reduce the effects of earthquakes like creating disaster-preparedness plan, identifying buildings weaknesses and trying to fix it, creating disaster kits, identifying hazards in your area and fixing them etc. [2] But there is not any model which can predict the level of damage done to buildings during earthquakes. The model which is built in this Project will help in reducing the destruction and loss of human lives which is resulting from aftershocks. The damage to a building is categorized into five grades. Each grade depicts the amount of damage caused to buildings after an earthquake.

II. LITERATURE REVIEW

System gaining knowledge has been broadly used for making earthquake predictions because of their potential to enhance over the years. With the big quantity of earthquake instrumental records, system learning strategies are capable sufficient to enhance performance and accuracy in earthquake prediction. [3] Multiple gadget gaining knowledge of methods consisting of, synthetic Neural network (ANN), Support Vector Machine (SVM), k-nearest neighbour (KNN), Naive Bayes (NB) and random forest algorithms have been exercised for earthquake prediction.

Many researches have been made at the damage price of homes in Japan. Mononobe [1970] become the primary to research them the overturning of tomb stones which is equivalent to ground seismic coefficient suggests the family members acquired from the harm information of numerous large earthquakes in Japan and theoretical curves expressed with the aid of everyday chance distribution feature. [4] Shiga [1976] discussed the harm rate of concrete homes the ones which are reinforced primarily based on his studies on the earthquake resistance capacity of current those which might be strengthened homes. He derived a simple formula to estimate roughly the resistance capability of lower bolstered concrete homes the use of the amount of column areas and wall regions through the evaluation of damaged in addition to undamaged homes inside the (1968) Tokachioki earthquake and the (1978) Miyagikenoki earthquake. He estimated the probability distribution of the earthquake resistance capacity of present low-rise strengthened concrete buildings and modeled the chance distribution by means of Gamma distribution, based totally on which he gave the relation between the damage charge of bolstered concrete buildings and the ground motion intensiy.

Wenrui Li, Nakshatra, Nishita Narvekar, Nitisha Raut, Birsen Sirkeci, Jerry Gao introduce us to the idea that a strong earthquake is accompanied by means of aftershocks. We will locate vicinity of these aftershocks via evaluation of arrival time of P-waves and S-waves. Statistics series from sixteen earthquake stations in SAC document format, which includes time series statistics and is a waveform, utilized by authors to examine trends in P-wave and S-wave. Information is clipped followed by noise removal to simplest acquire needed waveform by means of triggering set of rules and filters. [5] AR picker algorithm used to determine values of P-wave and S-wave arrival time which can be treated as extracted function. Waveform is then converted into ASCII layout. Statistics is then fed to extraordinary gadget getting to know models-SVM, decision tree, Random forest and linear regression for evaluation reason. Random forest distinguishes among earthquake leading and non-earthquake main statistics the pleasant, with an accuracy of 90. Use of triangulation technique is used to calculate epicenter, expect arrival time of P-wave and S-wave and the distinction among the two arrivals.

Earthquake activity is presumed as a spontaneous phenomenon that can damage huge number of lives and properties, and currently there is no any model exists that can predict the exact position, magnitude, frequency and time of an earthquake. Researchers have conducted several experiments on earthquake events and forecasts, leading to a variety of findings based on the factors considered. [6] The well-known Gutenberg and Richter statistical model found a correlation between the magnitude of earthquake and frequency of earthquake. For structural design, this earthquake probability distribution model was used. In supervision of the California Geological Survey, Petersen conducted research and suggested a model that is time-independent. This time independent model demonstrates that chances of occurrence of earthquake follow the Poisson's distribution model. Shen suggested a probabilistic earthquake forecasting model based on the strain studied between the behaviour of tectonic plates. Based on this model, higher measured strain results in a higher risk of earthquake.

III. PROPOSED WORK

The proposed system of this research work consists of 6 modules which are shown in flowchart diagram:

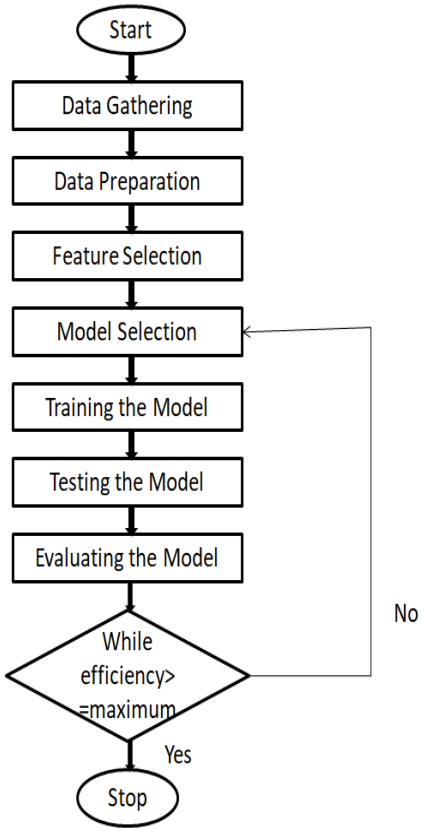

Fig1: Flowchart Diagram

A. Data Gathering

It is the process of gathering data from various sources like files, databases, etc. It is the first step of machine learning. The nature of predictive model depends on the quality and quantity of the collected data. The dataset is collected from hackerearth website. There are four excel files in dataset: Building_Structure.csv ,Building_Ownership_Use.csv , Train.csv & Test.csv file and one readme file which explains all the features of dataset. The gathered information can't be used immediately for analysis system because it contains missing data, unorganized text values, etc.

B. Data Preparation

It is the second step of machine learning. Machine learning models are built more accurately with the help of this step. The record that is accumulated from different assets is in a raw layout, so it is not possible to do analysis easily. So there are some steps done for converting the raw data into small dataset. This technique is known as data pre-processing.
The pre-processing techniques used for data preparation are:

- **Conversion of the data:** The specific and ordinal records is to be transformed into numeric values due to the fact machine gaining knowledge of models can most effective take care of numeric values.
- **Ignoring the missing values:** If the dataset incorporates lacking statistics, we will cast off that row or column which includes missing values. It's far an efficient technique however if there is a lot of missing records inside the dataset then it should not be finished.
- **Filling the missing value**: We are able to fill the lacking values manually through selecting imply, median, and many others.
- **Machine learning:** We can also make predictions with the help of the use of the present records and fill the lacking values within the dataset.
- **Outliers detection:** There can be a few error facts present in the dataset which diverge our interest from different values inside the dataset.

C. Feature Selection

It is the process of selecting some features manually or automatically which helps in predicting the output. The accuracy of the models is decreased if it contains irrelevant features. It is the most important concept of machine learning that affect the performance of the model. It reduces the redundant information and hence decisions may be made faster, improves the accuracy of the model and reduces the complexity of algorithm hence it saves time.

D. Model Selection

It is the process of selecting between the different machine learning, algorithms, e.g.: SVM, Random Forest, Lightgbm, XGboost, Neural Network, Decision Tree, etc. The model we choose should be simple and easy to explain, accurate, fast and should be scalable.

The dataset of this Project is labeled correctly, so it is a part of supervised learning. Classification is a category of supervised learning. The most used Classification algorithms are:

- K-Nearest Neighbor
- Decision Trees
- Random Forest
- Support Vector Machine
- XGBoost
- Neural Network
- Lightgbm

E. Training The Model

Training is the maximum important step in machine learning. In this, the organized facts are handed to system learning model to discover styles and make predictions. It outcomes in the version gaining knowledge of from the facts in order that it can accomplish the mission set. Through the years, with training, the model receives better at predicting. For training the classifier we use the training dataset. It is a dataset which is used to train a model. In this process, we pick specific features from the training set and then incorporate it into the model.

F. Evaluating The Model

After training your version, you have to test to see how it's acting. This is executed by way of testing the overall performance of the version on previously unseen data. If testing became performed at the same information that is used for training, you will no longer get an accurate measure, because the version is already used to the information, and finds the same styles in it, as it previously did. This will provide you with disproportionately excessive accuracy. When used on trying out records, you get a correct measure of ways your model will carry out and its pace.

IV. DFD (DATA FLOW DIAGRAM)

A. 0-Level DFD

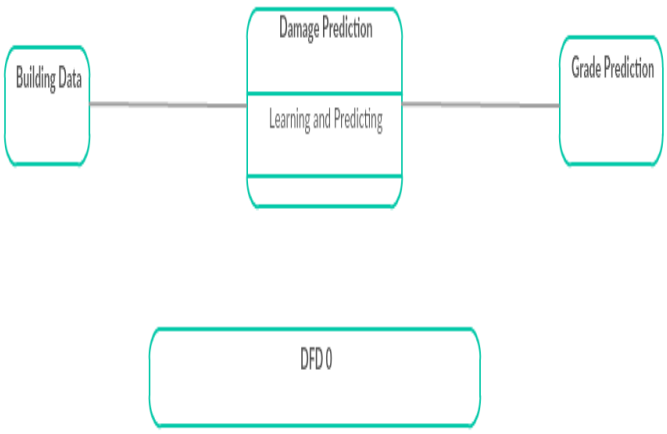

Fig 2: 0-Level DFD

- In the first step, we are provided the building data in the dataset.
- In the second step, we apply various machine learning algorithms like Random Forest, Xgboost, Neural Network etc. on this dataset to predict the result of our Project.
- In the third step after applying algorithms on dataset, we can predict the grade of buildings and thus estimate how much damage can be caused to it due to earthquake.

B. 1-Level DFD

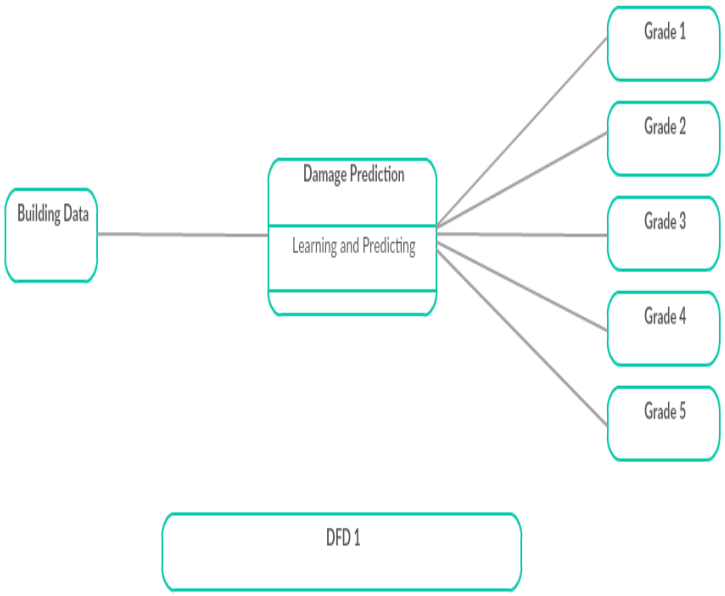

Fig 3: 1-Level DFD

In this 1-Level DFD we can further classify the prediction of grade into five categories. They are:

- **Grade 1(Minor) =>** There is light or no damage to walls of the building
- **Grade 2(Medium)** => There is flexural cracks on non-structural wall
- **Grade 3(Major)** => There is damage on buildings due to accidental affects
- **Grade 4(Partial collapse)** => There is some part of the building which is collapsed
- **Grade 5(Total collapse)** => There is significant damage to shear walls and entire building is collapsed.

V. ARCHITECTURE DIAGRAM

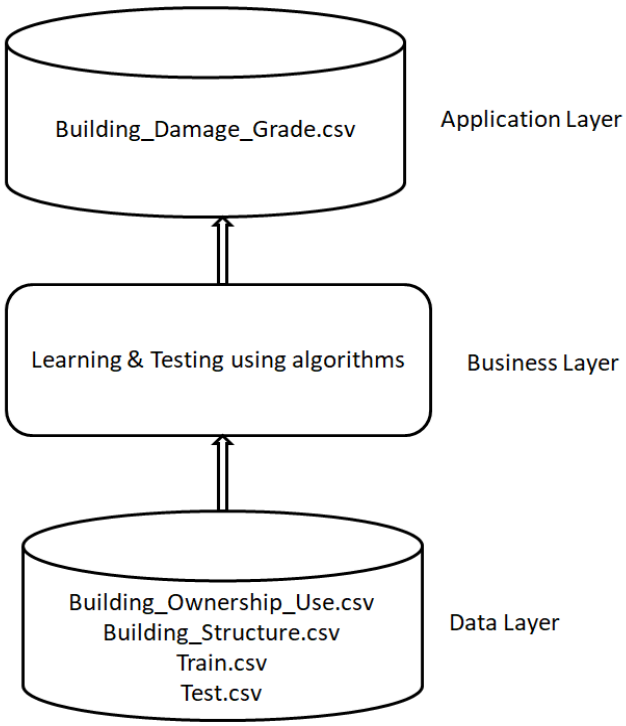

Fig 4: Architecture diagram

The architecture diagram consists of three layers:

A. Data Layer: It is the layer which contains the original or raw data. In this Project, the dataset consists of four csv files: Building_Ownership_Use data, Building_Structure data, Training data and Testing data. The data layer specifies the data in the dataset folder.

B. Business Layer: It is a layer where feature selection, model selection, training, testing and assessment of the model overall performance is achieved. We use machine learning algorithms for this technique. A few machine learning algorithms are Random Forest, Decision Tree, Ensembling, Lightgbm, Neural Network, XGBoost etc. After choosing the model, we carry out training and testing of the model. The dataset is divided into training set, test set. In training set, we teach the model and pick some unique functions which can be integrated into the model. Test dataset estimates how the model makes the predictions within the test dataset.

C. Application Layer: It is the layer which contains the predicted data. The final predicted output for grade estimation of buildings is stored in a Building_Damage_Grade.csv file. The application layer shows the final result of the machine learning problem. The grade estimation helps to predict the damage which can be caused to buildings during earthquakes. Therefore, necessary preventions can be taken to rescue building from earthquakes.

VI. CONCLUSION

Scientists have never anticipated a first-rate earthquake. Earthquake prediction should define 3 elements: the date and time, location and importance. Some Scientists can detect earthquakes after the earthquake is begun and they can send warnings to the areas before it reach there. This Project helps in reducing the effects of damage and destruction caused by aftershocks of earthquake. If we know the grade of buildings, we can take measures to reduce its weakness and try to strengthen it to face earthquake. The lower the grade of building, the more damage it is to be caused by earthquake. The higher the grade of building then it makes it easy for the building to survive an earthquake. The predicted output of buildings data helps in identifying safe and unsafe buildings. Using machine learning has helped us to make earthquake less painful and severe. It is a feasible option to prevent the damage to buildings and loss of human lives.

VII. ACKNOWLEDGEMENTS

I would like to express my very great appreciation to Mrs. Vanita Buradkar for her valuable and constructive suggestions during the planning and development of this research work. Her willingness to give her time so generously has been very much appreciated.

VIII. REFERENCES

[1] Rabia Tehseen, Muhammad Shoaib Farooq and Adnan Abid. "Earthquake prediction using expert systems: a systematic mapping study", Department of Computer Science, University of Management and Technology, Lahore 54770, Pakistan.

[2] Dr. S. Anbu Kumar, Abhay Kumar, Aditya Dhanraj, Ashish Thakur. "Earthquake prediction using machine learning", International Research Journal of Engineering and Technology (IRJET) Volume: 08.

[3] Rabia Tehseen, Muhammad Shoaib Farooq and Adnan Abid. "Earthquake prediction using expert systems: a systematic mapping study", Department of Computer Science, University of Management and Technology, Lahore 54770, Pakistan.

[4] Akenori Shibata, Masato Motosaka and Yoshinori Hagiwara. "Prediction of damage rate of building groups in urban areas by strong earthquakes", 12th World Conference on Earthquake Engineering, Auckland, New Zealand.

[5] 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.

[6] Dr. S. Anbu Kumar, Abhay Kumar, Aditya Dhanraj, Ashish Thakur. "Earthquake Prediction using Machine Learning", International Research Journal of Engineering and Technology (IRJET) Volume: 08.