```
In [1]: import pandas as pd

In [2]: import numpy as np

In [3]: pd.options.display.max_rows = 6

In [4]: pd.options.display.max_columns = 6

In [5]: index = pd.DatetimeIndex(start='20010101',freq='D',periods=10)

In [6]: pd.DataFrame(np.arange(10*10).reshape((10,10)),index=index)
Out[6]:
             0   1   2   3   4   5
2001-01-01   0   1   2   3   4   5 ...
2001-01-02  10  11  12  13  14  15 ...
2001-01-03  20  21  22  23  24  25 ...
2001-01-04  30  31  32  33  34  35 ...
2001-01-05  40  41  42  43  44  45 ...
2001-01-06  50  51  52  53  54  55 ...
            ... ... ... ... ... ...

[10 rows x 10 columns]
```

Fig. 4: sample output for jupiter Application

The python library matplotlib, pandas, os are used in our project. The detailed explation of each library is followed by

1) pandas— pandas library is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language. In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.
2) Matplotlib---- Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Create. Develop publication quality plots with just a few lines of code. Use interactive figures that can zoom, pan update.
3) Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+
4) numpy— NumPy is the fundamental package for scientific computing in Python NumPy arrays facilitate advanced mathematical and other types of operations on large numbers of data. Typically, such operations are executed more efficiently and with less code than is possible using Python's built-in sequences
5) pycharm— This is a Professional feature download PyCharm Professional to try. In this tutorial, you operate in Scientific Mode and use Matplotlib and NumPy packages to run and debug a Python code with data visualization. Before you start, ensure the following is installed Conda interpreter. Matplotlib package
6) OS— The OS module in Python provides a way of using operating system dependent functionality. The functions that the OS module provides allows you to interface with the underlying operating system that Python is running on – be that Windows, Mac or Linux.

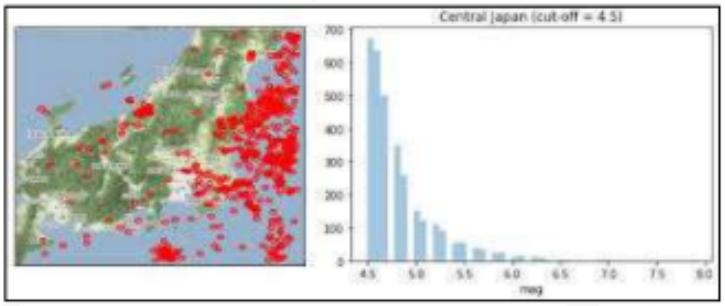

Fig. 5: sample output for matplotlib

The Matplot library used to find the targeted graphical output using numpy.and derive a data and clear output in user.

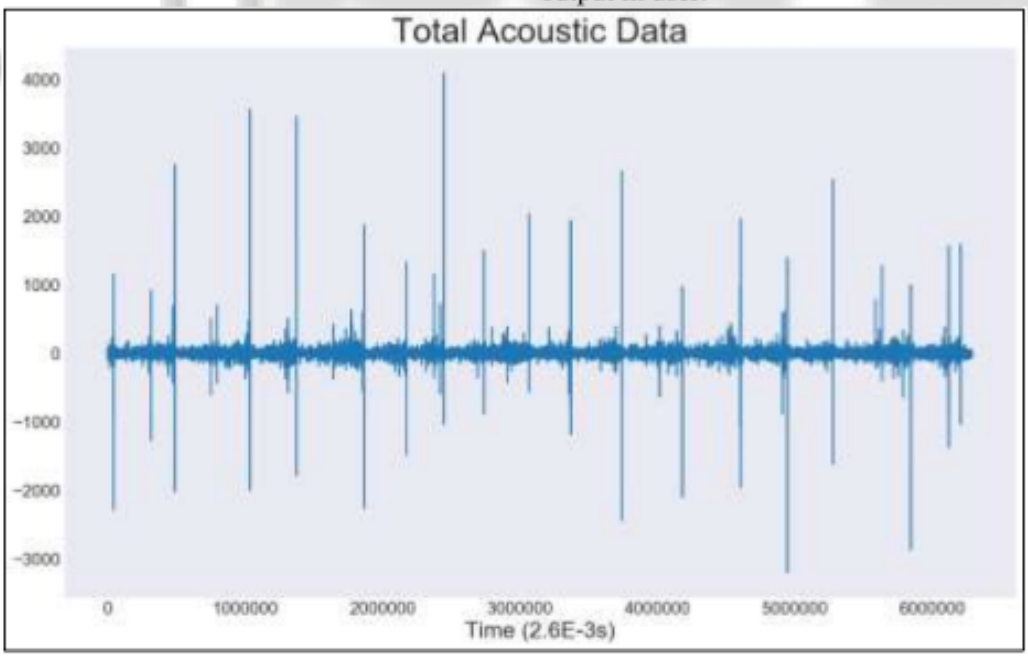

Fig. 6: sample output for pandas library

The library used to evaluvate a data and determind and formatting a datasets.it provided a statical represented a output.

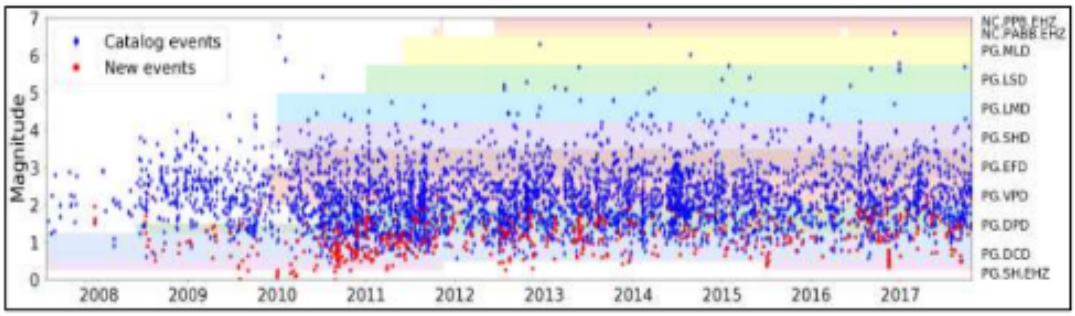

Fig. 7: Sample output of the mangnitude in matplotlib

Furthermore, the analysis of earthquake prediction results is carried. Past earthquake magnitude data are used as an input for the network.

V. Conclusion

This work presents that the Random Forest Classifier algorithm has the highest accuracy in predicting the damage due to earthquakes, based on the F1 score calculated for each of the four algorithms previously mentioned in this work. K-Nearest Neighbors has been observed to be the second most preferred algorithm for earthquake damage prediction. On analysis of the materials that help curb damage to buildings during an earthquake, the work concludes that Reinforced Concrete is the material most suited to the cause. Earthquakes are well known to excite electromagnetic pulse, that cause tremors under the Earth's crust. These electromagnetic pulses are shielded effectively by Reinforced Concrete. Reinforced concrete has a low tensile strength, and hence Steel bars are used, which are embedded in the concrete sets. This provides Reinforced Concrete with immense ability to withstand natural calamities such as Earthquakes. The applications of this work can be further extended earthquaketo predict damage caused by Earthquakes in areas & time also possiblefor which a similar and relevant dataset can be obtained.

Reference

[1] C.P Shabariram and K.E Kannammal "Earthquake prediction using map reduce framework" 2017 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–6, 2017.

[2] You-Po Su and Qing-Jie Zhu "Application of ANN to Prediction of Earthquake Influence" 2009 Second International Conference on Information and Computing Science, vol. 2 pp. 234-237, 2009

[3] Cao Li and Xiaoyu Liu, "An improved PSO-BP neural network and its application to earthquake prediction," in 2016 Chinese Control and Decision Conference 2016, pp. 3434–3438

[4] Xueli Wei, Xiaofeng Cui, Chun Jiang, Xinbo Zhou "The Earthquake Probability prediction based on weighted factor coefficients of principal components "2009 Fifth International Conference on Natural Computation, vol. 2, p. 608-612, 2009

[5] Dezhang Sun, Baitao Sun "Rapid prediction of

[6] Jianwen Zhang, Changyong Wang "Sheer Capacity computation of steel reinforced lightweight concrete beams" International Conference on Mechanic Automation and Control Engineering, 2010. p 1502-1506

[7] The Hackerearth platform official website "https://www.hackerearth.com"

[8] Python Seaborn Library pydata official website "https://seaborn.pydata.org/generated/seaborn.scatterplot.htm"

while that of the train dataset was 617,725. The training data and the test data percentages has been observed to be 70.50% and 50.50% (approximated to80% and 60% respectively).

III. PROPOSED SYSTEM

In future, same Mapper and Reducer class implemented with pandas and matplatlib frame work components working effactive manner. Pandas handale with a data is esay that's way proposed system run with effactively. visuvalation method is matplatlip framework working with overcome the exsisting system. data parsing &data format conversion is esay way to possible in pandas. Given query and get final result in graphically. The graphs compare the number of affected buildings (count) for a particular Damage Grade to their corresponding foundation type, roof type and ground floor type respectively. The ratios indicate the likelihood of buildings with the given material's ability to sustain damage against earthquakes. The ratios are considered here instead of directly comparing the number of affected buildings as the former can take into account the variation observed among the Damage Grades in each case, irrespective of the number of affected buildings . A lot of research is being done in this field, as a result of which manifold approaches to predict damage on account of earthquakes have been worked upon and developed.

IV. DESIGN AND IMPLEMENTATION

This covers the technique and flow of events that were to perform the prediction process. The pred methodology itself is composed of three integral steps preprocessing, model selection and the final pred process. In the dataset, a building was uniquely identif 4 attributes: Building Identification, District Identific Municipality Identification, Ward Identification. attributes were added to the training data for identifyi building damage grade. There were 33,417 entries attribute pertaining to whether building repairs on earth affected buildings had started or not, that were found blank. Based on the assumption that since there w formally documented record of the commencement repairs, the blank values were assumed to be not rep Such filling was done on the basis of the worst case sce possible to get optimal results.

The models developed by the individual algor were trained on the training dataset and then test dat used for final prediction of damage grades and for evalu Since on evaluation, Random Forest Classifier algorith found to possess the highest F1 score, the mode considered for the prediction process.

On dropping the following Boolean attributes the dataset we found an increase in the model score 0.75127 to 0.76503 – whether the building has seconda as an institution, whether the building has other geotec risks.

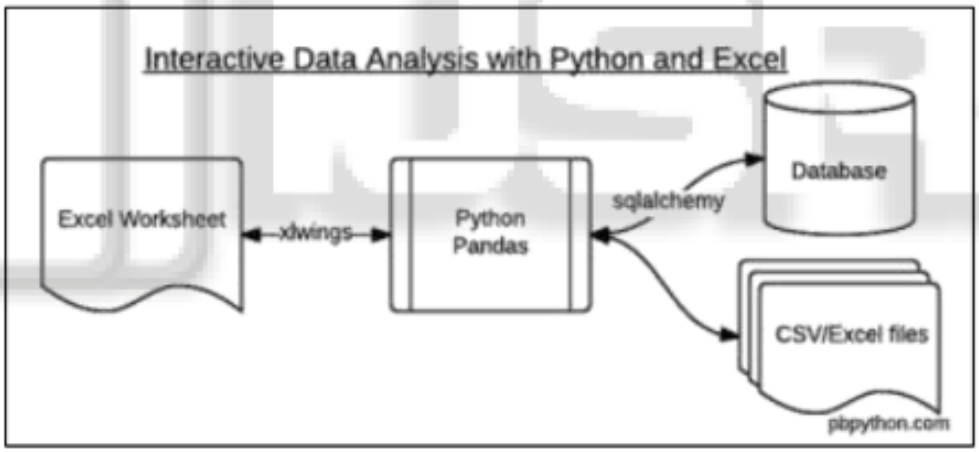

Fig. 2. Block Diagram of pandas and datasets

The fig.2 is the block diagram of the Web Reader Application. This diagram shows the flow of the Web Reader Application. The user can give the prefixed URL input to the Web Reader Application. And the python libraries like requests, bs4, textblob are processed the URL and retrie specified web content from the given particular URL. A retrieved web content is converted into the audio format in native language Tamil.

```
In [65]: print type(data.iloc[100])          # result of type series because only one row selected
         print type(data.iloc[[100]])        # result of type DataFrame because list selection used
         print type(data.iloc[2:10])         # result of type dataframe since there are two rows selected
         print type(data.iloc[1:2, 3])       # Series result because only one column selected
         print type(data.iloc[1:2, [3]])     # DataFrame result with one column be only one column selected
         print type(data.iloc[1:2, 3:6])     # DataFrame results because multiple columns and multiple rows.

         <class 'pandas.core.series.Series'>
         <class 'pandas.core.frame.DataFrame'>
         <class 'pandas.core.frame.DataFrame'>
         <class 'pandas.core.series.Series'>
         <class 'pandas.core.frame.DataFrame'>
         <class 'pandas.core.frame.DataFrame'>
```

Fig. 3: sample output for jupiter Application

Abstract— An earthquake is shaking of the surface of the Earth, which caused as the result of movable plate boundary interactions .Earthquakes are measured using remarks from seismometers with Richter magnitude scale. Ground rupture, Landslides, Soil liquefaction and Tsunami are the main effects created by earthquakes. Today's earthquake warning systems used to provide regional notification of an earthquake in progress. Many methods have been already developed for predicting the time and place in which earthquakes will occur, but it did not predicted using big data analytics. This journal know that the Standard Deviation to identify next earthquake happening from tons of international geological survey data using data analysis in pandas & matplotlib framework. It's the top-level component of all the ones that you will consider in the following point current location shakes per minute. Other than above mentioned features separate pandas and matplot function implemented to analyze sheer number of earthquakes per day. Final result shows which location suffered from maximum number of shakes and priority of earthquake occurrence location and Time.

Keywords: Earthquake, Seismic waves, Seismometer, Richter magnitude scale, Tsunami, Earthquake warning systems, Big Data, Map Reduce

I. INTRODUCTION

Today, big data analytics is one of the most booming markets. When Google search engine launched image search feature, it had indexed more than 300 million images. In every minute so many video content are uploaded in YouTube update their wall in every minute. Search engines logging 600 million queries daily. There are different data centers where people can store vast amount of data, such as IBM Server, EMC Server etc. On the other hand AWS (Amazon Web Services) provide a host of services to store, process and analyze the data at scale in a cost effective manner. Big data term refers collection of large datasets that are distributed, multi-dimensional and complex that it becomes difficult to processing on hand traditional data processing applications.

II. EXISTING SYSTEM

The existing system addresses novel methodology to predict next earthquake. Apache hadoop is designed to run in a distributed environment and it manages the collection of various nodes running map and reduce function. In this system data analysis performed on earthquake data in year wise and location wise. The result indicates that next possible earthquake location identified correctly. The exisiting system Mapper and Reducer class applicable any U.S. geological earthquake survey data.in this funcation some drawbacks in data handaled but pandas overcome the problems thats way using pands & matplotlib frame work reduse the problem and predict the possible output.

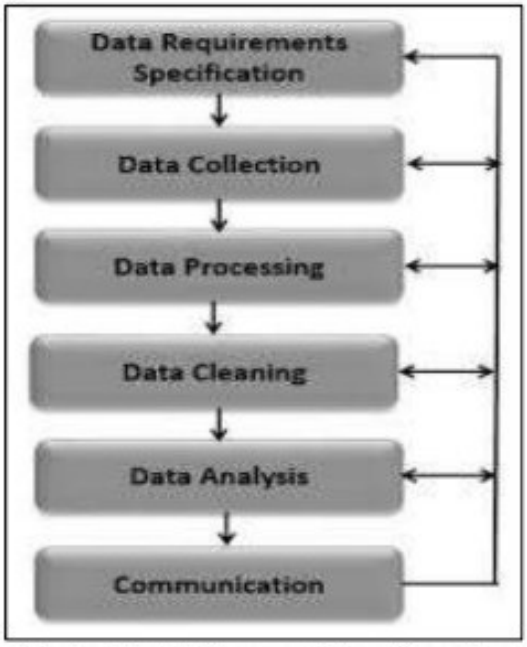

Fig. 1: Block Diagram of flow funcation

Earthquakes have always caused incalculable 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 post an earthquake. It can help identify safe and unsafe buildings which helps to predict damage prone areas and thus avoiding death and injuries resulting from the aftershock of 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 of damage. Models of this kind can help save as many lives as quickly as possible and turn out to be an efficient and cost-effective solution. It can be further improved by the inclusion of distribution of resources like food, clothes, medical, monetary supplies based on the extent of human casualties and the damage incurred by the various structures.

The main *disadvantage* of the existing system is highly expensive. And the portability of the Device is very difficult. The Existing system using only converts using modules county and number of earth quakes only provided. The dataset was found to have a total of 1,038,900 records, with the number of records in the test dataset being 421,175,