Course : Artificial Inteligence

Project: Earthquake Prediction Model Using Python

Document : Phase 2 Submit


I. INTRODUCTION

Today, big data analytics is one of the most booming markets.When Google search engine launched image searchfeature,it had indexed more than 300 million images. In every minute so many video content are uploaded inYouTube update their wall in every minute. Search engines logging 600 millionqueries 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 referscollection of large datasets that are distributed, multidimensional and complex that it becomes difficult toprocessing on hand traditional data processing applications

II. EXISTING SYSTEM

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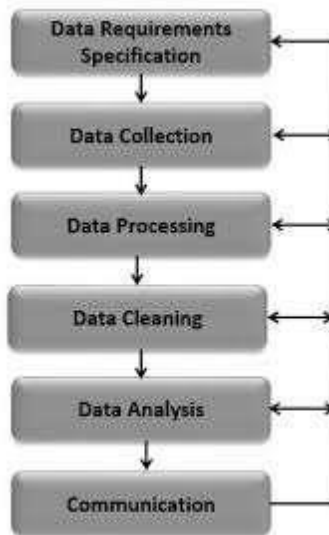
Block Diagram of flow function

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, 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 used to perform the prediction process. The prediction methodology itself is composed of three integral steps: data preprocessing, model selection and the final prediction process. In the dataset, a building was uniquely identified by 4 attributes: Building Identification, District Identification, Municipality Identification, Ward Identification. These attributes were added to the training data for identifying the building damage grade. There were 33,417 entries in the attribute pertaining to whether building repairs on earthquake affected buildings had started or not, that were found to be blank. Based on the assumption that since there was no formally documented record of the commencement of the repairs, the blank values were assumed to be not repaired. Such filling was done on the basis of the worst case scenarios possible to get optimal results. On dropping the following Boolean attributes from the dataset we found an increase in the model score from 0.75127 to 0.76503 – whether the building has secondary use as an institution, whether the building has other geotechnical risks.

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 objectoriented 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.
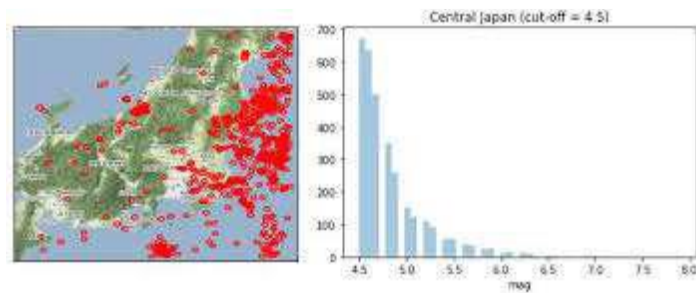
The Matplot library used to find the targeted graphical output using numpy.and derive a data and clear output in user.
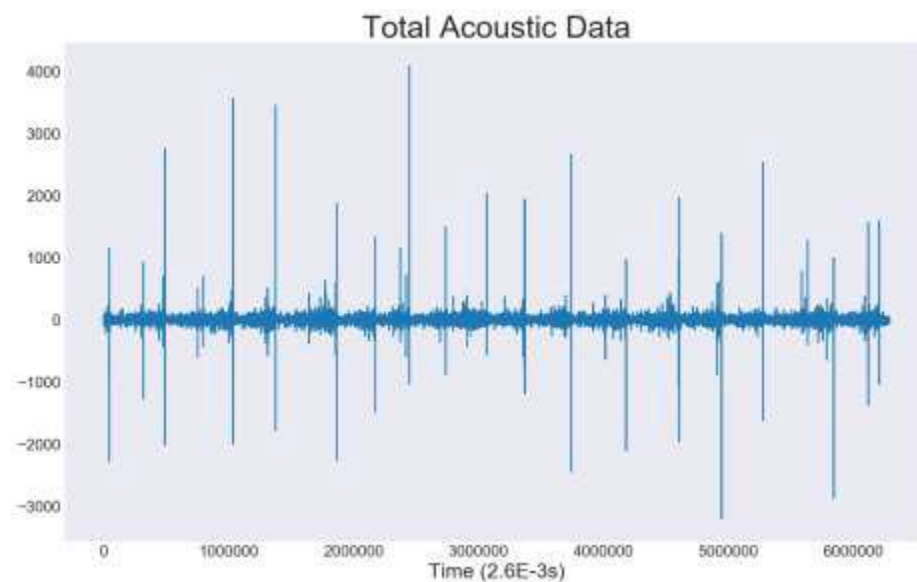
The library used to evaluvate a data and determind and formatting a datasets.it provided a statical represented a output.
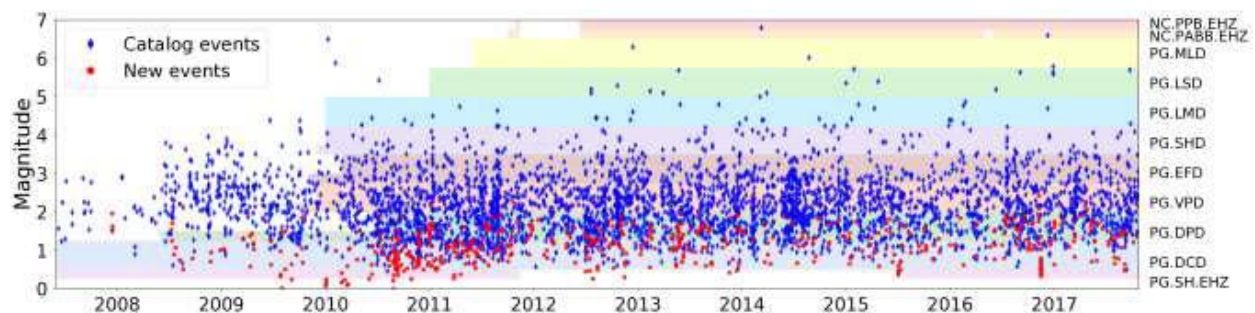
Furthermore, the analysis of earthquake prediction results is carried. Past earthquake magnitude data are used as an input for the network.

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. KNearest 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