Earthquake Prediction

Mini-Project report submitted in partial fulfilment

of the requirements of the degree of

Bachelor of Engineering

by

Digole Suhas Sunil EU1162095

Mendapara Harsh Gordhan EU1162099

Thakur Manthan Uday EU1162006

Under the guidance of

Mr. Chetan Patil

Department of Computer Engineering

St. John College of Engineering and Management

University of Mumbai

2018-2019

CERTIFICATE

This is to certify that the project entitled **"Earthquake Prediction"** is a Mini-Project report of

"Digole Suhas Sunil" (EU1162095)

"Mendapara Harsh Gordhan" (EU1162099)

"Thakur Manthan Uday" (EU1162006)

submitted in partial fulfilment of **"Bachelor of Engineering"** in **Computer Engineering"** as laid down by University of Mumbai during the academic year 2018-2019.

Mr.Chetan Patil
Guide

Dr. G. A. Walikar
Head of Department

Dr. G. V. Mulgund
Principal

Mini-Project Report Approval for T. E.

This project synopsis entitled ***Earthquake Prediction*** by ***Digole Suhas Sunil, Mendapara Harsh Gordhan, Thakur Manthan Uday*** is approved for the degree of ***Bachelor of Engineering*** in ***Computer Engineering*** from ***University of Mumbai***.

Examiners

1.--

2.--

Date:

Place:

Abstract

Earthquake is the major problem in this world from all the disasters. The waves with high energy are able to demolish different types of infrastructures. To the upcoming problem from many years, data analysis and deep learning is well developed with the help of different kinds of prediction techniques earthquake analysis and its prediction is carried out. Our project takes the significant earthquake data from different regions and can visualize in different types of graphs, plotting the data into world map and predict the magnitude by of the earthquake along with locations the simple neural network. Our project can help the government/authorities to protect and minimize the loss of life and property.

Keywords:- Disaster, Neural Network, Analysis , Prediction , Visualization.

Contents

List of Figures

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)
Digole Suhas Sunil (EU1162095)

(Signature)
Mendapara Harsh Gordhan (EU1162099)

(Signature)
Thakur Manthan Uday (EU1162006)

Date:

Chapter 1

Introduction

An earthquake (also known as a quake, tremor or temblor) is the shaking of the surface of the Earth, resulting from the sudden release of energy in the Earth's lithosphere that creates seismic waves. Earthquakes can range in size from those that are so weak that they cannot be felt to those violent enough to toss people around and destroy whole cities. The seismicity, or seismic activity, of an area is the frequency, type and size of earthquakes experienced over a period of time. At the Earth's surface, earthquakes manifest themselves by shaking and displacing or disrupting the ground. When the epicentre of a large earthquake is located offshore, the seabed may be displaced sufficiently to cause a tsunami. Earthquakes can also trigger landslides, and occasionally volcanic activity.In its most general sense, the word *earthquake* is used to describe any seismic event—whether natural or caused by humans—that generates seismic waves. Earthquakes are caused mostly by rupture of geological faults, but also by other events such as volcanic activity, landslides, mine blasts, and nuclear tests. An earthquake's point of initial rupture is called its focus or hypocentre. The epicentre is the point at ground level directly above the hypocentre.Tectonic earthquakes occur anywhere in the earth where there is sufficient stored elastic strain energy to drive fracture propagation along a fault plane. The sides of a fault move past each other smoothly and a seismically only if there are no irregularities or asperities along the fault surface that increase the frictional resistance. Most fault surfaces do have such asperities and this leads to a form of stick-slip behaviour. Once the fault has locked, continued relative motion between the plates leads to increasing stress and therefore, stored strain energy in the volume around the fault surface. This continues until the stress has risen sufficiently to break through the asperity, suddenly allowing sliding over the locked portion of the fault, releasing the stored energy. This energy is released as a combination of radiated elastic strain seismic waves, frictional heating of the fault surface, and cracking of the rock, thus causing an earthquake. This process of gradual build-up of strain and stress punctuated by occasional sudden earthquake failure is referred to as the elastic-rebound theory. It is estimated that only 10 percent or less of an earthquake's total energy is radiated as seismic energy. Most of the earthquake's energy is used to power the earthquake fracture growth or is converted into heat generated by friction. Therefore, earthquakes lower the Earth's available elastic potential energy and raise its temperature, though these changes are negligible compared to the conductive and convective flow of heat out from the Earth's deep interior .One of the most devastating earthquakes in recorded history was the 1556 Shaanxi earthquake, which occurred on 23 January 1556 in Shaanxi province, China. More than 830,000 people died. The 1960 Chilean earthquake is the largest earthquake that caused the greatest loss of life, while powerful, were deadly because of their proximity to either heavily populated areas or the ocean, where earthquakes often create tsunamis that can devastate communities thousands of kilometres away. Regions most at risk for great loss of life include those where earthquakes are relatively rare but powerful, and poor regions with lax, unenforced, or non-existent seismic building codes.

1.1 Problem Statement

- Earthquakes are among the natural disasters created by sliding of the large land plates.
- It generally takes place due to the movement of the tectonic plates under the earth's surface.
- Here we choose the previously tested problem statement.
- Our problem statement will reflect the critical situation of the earthquake worldwide and how we can predict the earthquake and overcome its effects using the artificial neural network.
- In this case, we can show the maximum number of magnitude as well as the prediction with the help of graph and functions.

1.2 Solution Approach

- As we know that earthquake is one of the most destructive natural disasters in the world which may lead to other natural calamities which may cause a great loss of life and property.
- There must be a solution to this problem that how we can overcome this scenario by predicting the earthquake sensitive region, the magnitude of which the earthquake may take place and how we can minimize the loss of life and property.
- Thus we use an artificial neural network and a function relu to predict and visualize the situation.
- Here we consider the attributes such as date, latitude, longitude and depth as the input and give magnitude as the output.
- Here we will first analyze the data and convert it into a simplified form and use the prediction algorithm to get magnitude of the earthquake in the specified region and then at the last visualize it using graphs.

1.3 Objectives

- Our main objective is to show the magnitude-scale of the earthquake.
- To show our results in graphical manner.
- To predict the earthquake prone areas.
- By predicting the earthquake prone region we can minimize the loss of life and property.
- We may also be able to overcome the destructive situation in less time.
- With the advancement in the technology and with the help of artificial neural network and machine learning we can not only reduce but control the terrifying aftereffects of an earthquake.

1.4 Scope of the project

- Our project will be able to analyse and visualize the earthquake situation.
- With the help of which we will be able to control the situation.
- Our project can do analysis and prediction.
- Our project cannot do ground-motion analysis.

➢ The scope of our project follows the necessary step:

- Analysis
- Prediction
- Visualization

Chapter 2

Review of Literature

2.1 Introduction

Earthquake is the major problem in this world from all the disasters .The waves with high energy is use to demolish different types of infrastructures .To the upcoming problem from many years ,data analytics is well developed, with the help of different kinds of prediction techniques earthquake analysis is carried out. In this technical paper the earthquake analysis is carried out by applying map reduce function and statistical model is created. In this model large amount of datasets are been used. For processing a huge database after storing it Apache Hadoop tool is to be used. The various types of functions like map function, reduce function, etc. are stored in jar file and java programming is to be carried. Then job tracker is processed and gives output in form statistical graphs and pie-charts. Different types of results i.e. output is been display on basis of date, time and location. By using pie chart total analysis is to be carried out and maximum value is to be calculated. From Bar graph analysis as per location number of earthquakes occurred verses location their results are displayed.

2.2 Earthquake data analysis and visualization using Big –Data Tool

Earthquake is natural process which takes place due to the movement of tectonic plates below the surface of the earth. Earlier due to lack of knowledge and scientific instruments it was difficult for the scientists to find a solution to this problem, but due to advancement in science and technology these days earthquake can be predicted. Seismic waves travelling through earth surface carry a lot of data regarding the values buried under the earth crust. Seismic data processing has application in earthquakes and after shock detection. Big data technology analyse any type of data whether structured or unstructured. In this technique data can be analysed within few seconds regardles of the traditional techniques. An algorithm used to Analysis seismic data has two phases i.e. (mapper-phase and reducer-phase) where mapper phase is used for magnitude above 8 and reducer phase is used for earthquake having magnitude.

2.3 Magnitude of Earthquake prediction using neural network

Earthquake is natural occurrence which takes place due to the movement of tectonic plates below the surface of the earth. It creates an irretrievable financial and physical harm. Apparently the prediction of the earthquake is really vital, to our security. Various parameters such as latitude, magnitude, longitude, date, time, depth, temperature changes are taken into consideration. Many researchers work in earthquake prediction using geometric fields. In the recent years considerable work has been done about variations of earths geomagnetic fields for earthquake prediction. Despite of this earthquake prediction is a open problem now. However a neural network model can be trained, where the output of the neural network is the magnitude of the earthquake two days before its occurrence. This can be done by taking various attributes into consideration and calculating the magnitude of the earthquake using geometric fields declination and horizontal almost every time before the occurrence of the earthquake. Also the simulation results are promising.

Chapter 3

Requirement Analysis

In order to develop the specific system that will operate in the target environment and meet the specific objectives requirement analysis was conducted. Through observations of earthquake prediction with the research conducted by the other scientists in the same a lot of information was obtained.

3.1 METHODOLOGY

Agile development was adopted. This approach combines both the extreme programming XP and scrum methodologies to achieve the desired objectives the researcher aimed at tapping into the strength from the XP programming especially program refactoring and combining the scrum capabilities in order to overcome the different challenges found in embedded environment.

3.2 SYSTEM REQUIREMENTS

This defines how the user expectations will be met by the system, they are classified into:

3.2.1 FUNCTIONAL REQUIREMENTS

The system should be able to meet the following functionalities.

- Will be able to detect the magnitude of an earthquake in a specified region.
- Will perform the functions of analyzing, predicting and visualizing an earthquake prone region.

3.2.2 NON- FUNCTIONAL REQUIREMENTS

The system should be able to meet the following non-functionalities:

- Our project focuses on creating awareness of the dreadful situation caused due to an earthquake and how we can overcome it.
- Our project will help the government organization to plan for such natural calamities and reconstruct the affected area.
- It may not only help the suffered but also help in the development of our nation.

3.3 Minimum Hardware/Software Requirements

Hardware Requirements: - Laptop, Mouse, Keyboard, Internet Connection.

Software Requirements: - Jupyter Notebook, Sublime text editor, Spyder ide.

Chapter 4

Design

4.1 Use Case Diagram

A use case diagram is a behaviour or dynamic diagram of UML. Use cases are the set of services, functions, and actions that system needs to perform. A use case is a methodology which is used in the system analysis to clarify, identify, as well as organize system the requirements.

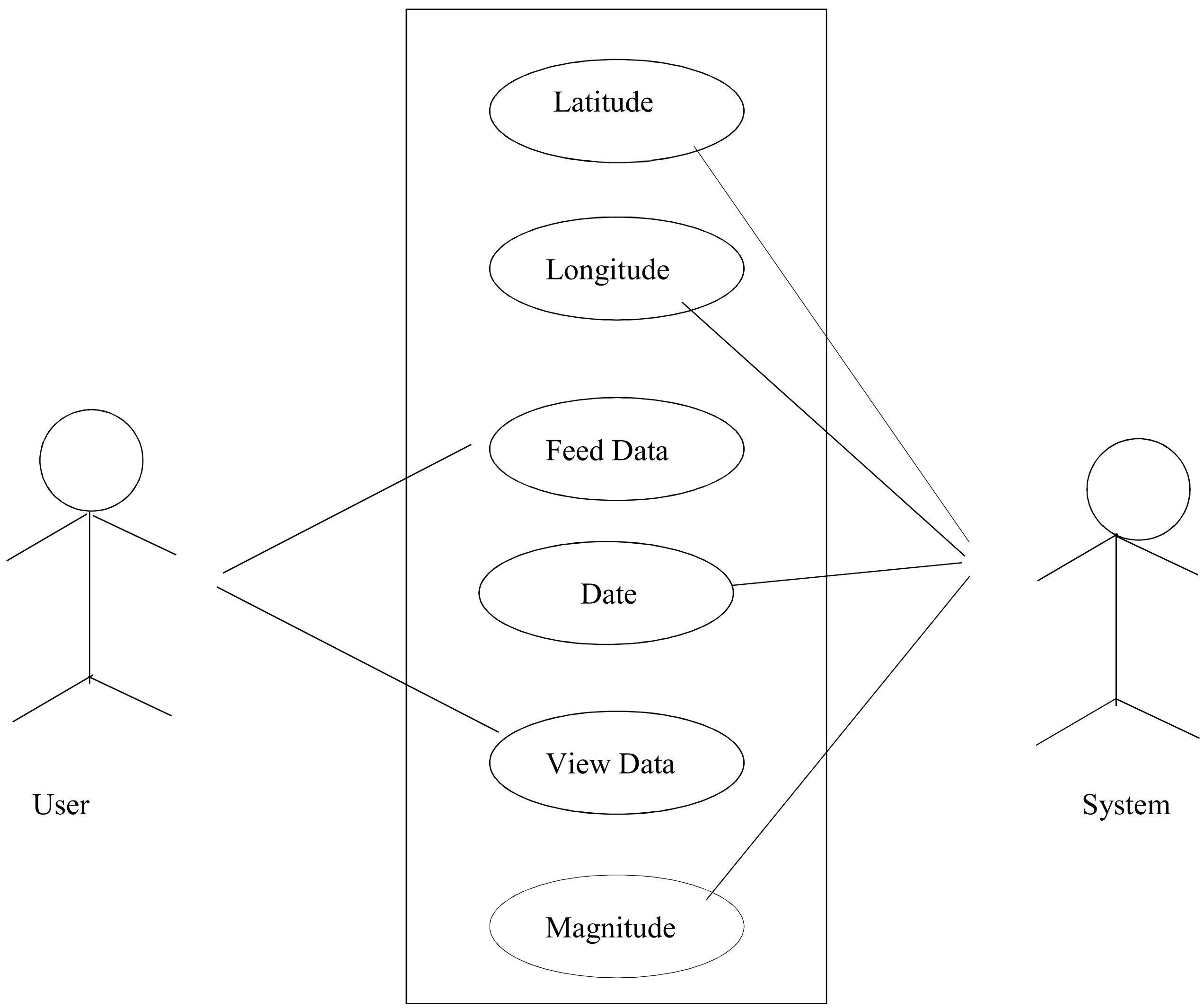


Fig 4.1: Use Case Diagram

4.2 DFD Diagram

A data flow diagram tell us that the data is processed by a system in the form of inputs and outputs. The name indicates how information flows, from where data comes, and how it gets stored. In DFD, the path for the data to move the information from one system to another system is shown. It is the graphical way of representation of the data flow in and out of the system. This simulation is to developed and understand the parameters of the natural disaster. Earthquake is a natural occurrence which helps us to visualize the different types of output and explain the flow of data.

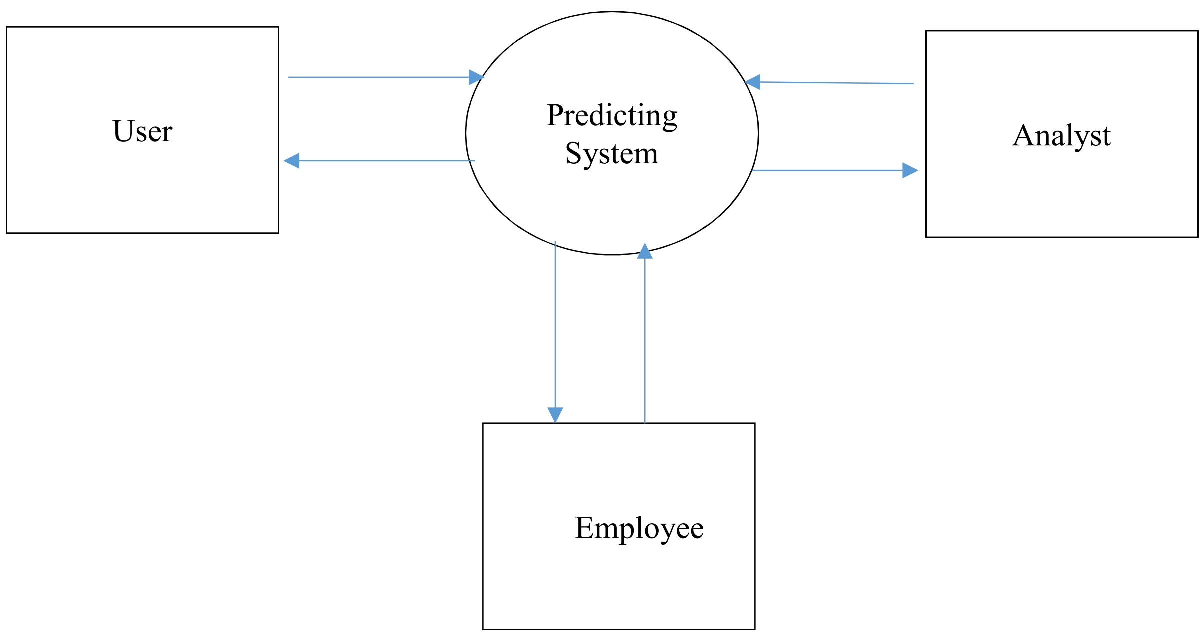


Fig 4.2 : DFD level 0

4.3 Class Diagram

Class Diagram is a static type diagram and is used for visualizing, documenting different facets of the system. It's main purpose is to make the static model view of an application. Class diagrams from UML phase are the most popular diagrams and are used for construction of a software application.

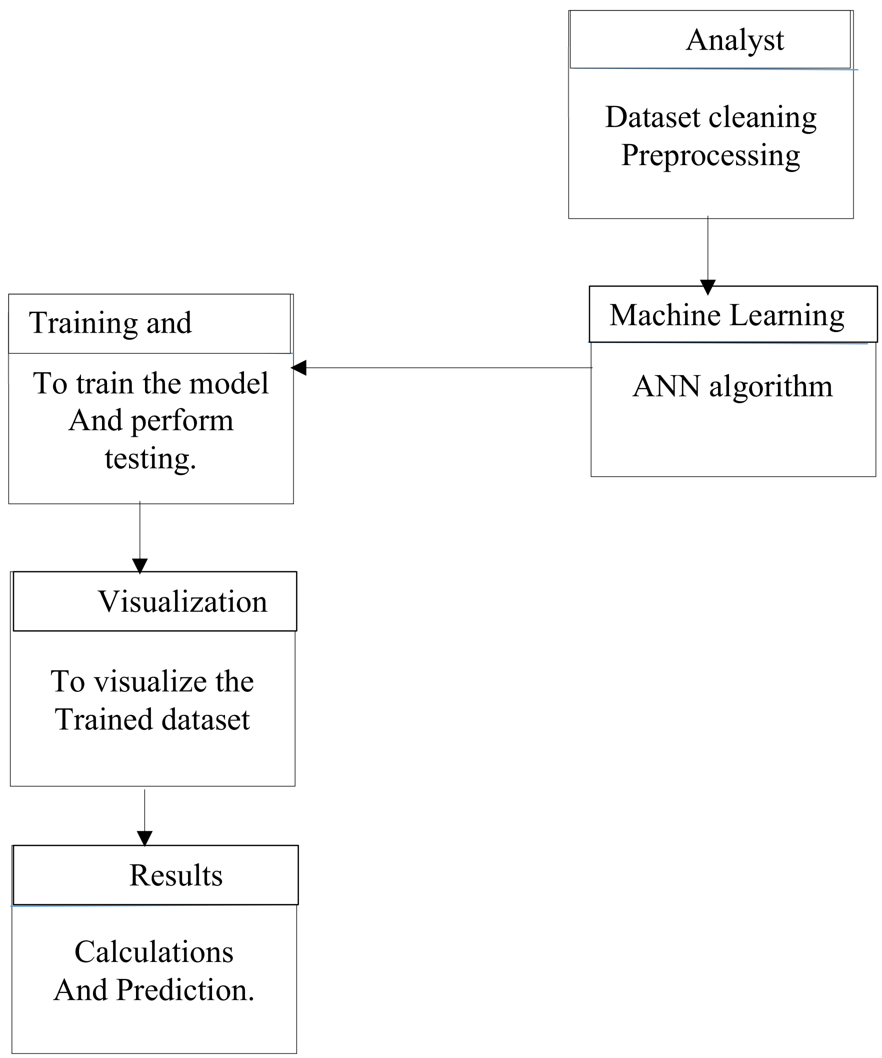


Fig 4.3: Class Diagram

4.4 Activity Diagram

They are used to show the flow of message from one activity to the other activity. Activity is the specific operation.

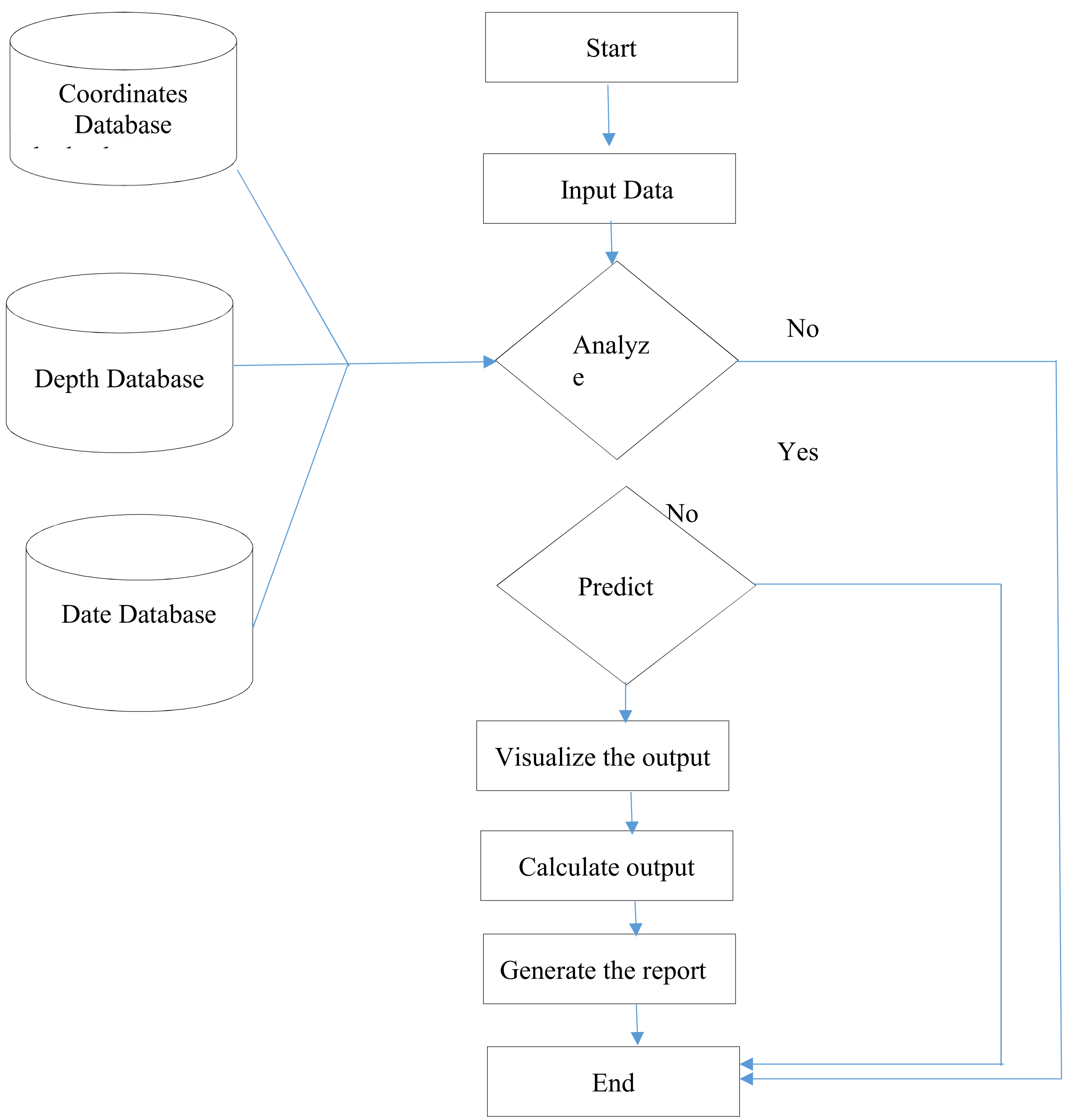


Fig 4.4: Activity Diagram

Chapter 5

Report of the project investigation

5.1 Proposed System

The proposed method for developing the system consists of mainly three major steps. Firstly, data is collected and sorted for relevancy from various sources. Then analysis is carried out on the collected data by understanding their attributes values, reducing errors by replacing null values, selecting specific features for taking input and output for earthquake prediction and finally we calculated statistical measure of each attributes. Secondly prediction is carried out on sorted dataset, here the obtained data is initially separated into training and testing purpose. A simple ANN is designed and a suitable algorithm yielding best accuracy is chosen to predict the magnitude value. At last the visualization of the calculated output with different attributes is to be carried out and illustrated in the graphical format.

5.1.1 Flow Chart of Proposed System

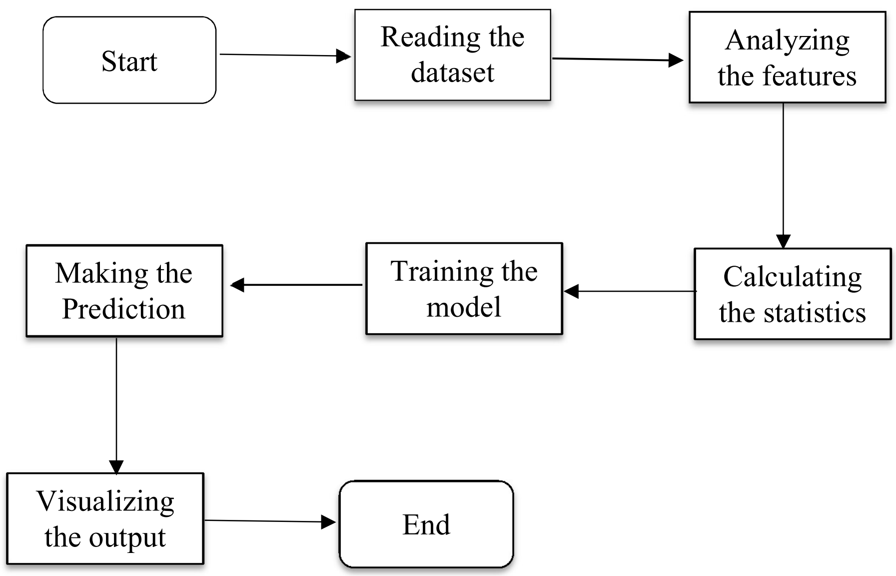


Fig 5.1.1 Flowchart of Proposed System

5.2 Data Sources

This project attempts to predict the magnitude value with respect to the previous value and trends. It requires historic data of significant earthquake as the project also emphasizes on data mining techniques, so it is necessary to have a trusted source having relevant and necessary data required for the prediction. We will be using Significant Earthquakes 1965-2016 website (https://www.usgs.gov/) as the primary source of data. This site contains all the details of different types of surveys based on natural disasters and new technologies which are being used for the researches. As this dataset contains the data of the earthquake with magnitude 5.5 or higher. There is no API provided by the website for providing data. We have written scripts to scrape all the required data from USGS website. Other sources include different types of significant data on earthquakes and relevant news related to natural disasters which might reflect the value and status of the earthquakes magnitude.

5.3 ANN Design and Training

The main problem in predicting magnitude from the significant earthquake data. There are some features which are fixed for prediction. We cannot draw any mathematical relations among the variables. There are no laws of predicting the magnitude using these variables.

For this kind of chaotic system the neural network approach is suitable because we do not have to understand the solution. This is the major advantage of neural network approaches. On the other hand in the traditional techniques we must understand the inputs, the algorithms and the output in detail. With the neural network we just need to simply show the correct output for the given inputs. With sufficient amount of training, the network will learn to ignore any inputs that don’t contribute to output.

In our purposed system, there is a training phase where some parameters name weights are found from this section and Backpropagation Algorithm is used for this training phase. These weights are used in prediction phase using same equations which are used in training phase.

5.3.1 Dataset Creation

First of all, a dataset is created for training the artificial neural network. The collected data are arranged according to the format for the library we use for training. The dataset should be of exact format that FANN specifies. It includes number of training pair, number of output in the first line of the dataset file and data from second line.

5.3.2 Dataset Normalization

The data is normalized before being input to the ANN. The input vectors of the training data are normalized such that all the features are zero-mean and the unit variance. The target values are normalized using min and max function such that all the values are converted into the values within the range of 0 to 1. The minimum value is represented by 0 and the maximum value is represented by 1.

equation no (1) $z = \frac{x - \min(x)}{\max(x) - \min(x)}$

5.3.3 Activation Function

We use ReLU (Rectified Linear Unit) function as the activation function at both hidden layer and output layer. The function and its derivative both are monotonic. But the issue is that all the negative values become zero immediately which decrease the ability of the model to fit or train from the data properly. That means any negative input given to the ReLU activation function turns the value into zero immediately in the graph, which in turns affects the resulting graph by not mapping the negative values appropriately. The formula is deceptively simple given below.

equation no (2) $R(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$

It avoids and rectifies vanishing gradient problem. ReLU is less computationally expensive that tanh and sigmoidal because it involves simple mathematical operations.

5.3.4 Backpropagation with Feedforward Neural Network

The main steps using the Backpropagation algorithm as follows:

Step 1: Feed the normalized input data sample, compute the corresponding output.

Step 2: Compute the error between the output(s) and the actual target(s).

Step 3: The connection weights and membership functions are adjusted.

Step 4: If Error > Tolerance THEN go to Step 1 ELSE stop.

5.3.5 Training Parameters

In order to select optimal parameters for the neural network, simulation is carried out. A model of a neural network is constructed and simulated. Test runs are carried out and the model yielding the best accuracy is selected for implementation. The best model so far has the following parameters:

Initial weights : 0.10
Learning rate : 0.001
Total Layers: 4
Input Neurons: 4
Hidden Layer Neurons: 4
Output Neurons: 1
Activation Function: ReLU
Limit of epochs: 50000
Minimum errors: 0.00001

5.3.6 Model Design

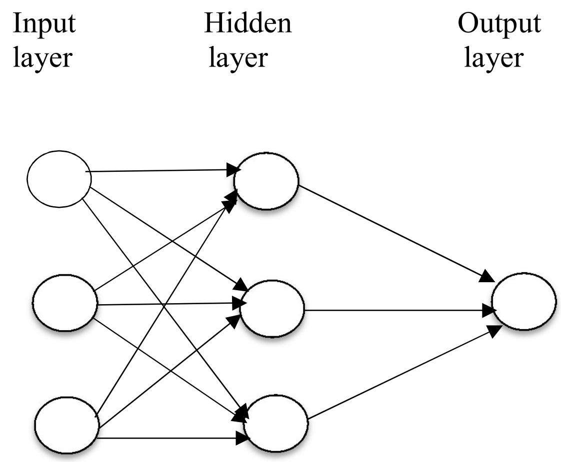


Fig 5.3.6 Neural Network Model

Chapter 6

Result and Analysis

6.1 Analysis

In this project, the factors are taken for data analysis of significant earthquake data are: Date, Time, Latitude, Longitude, Depth, Depth Error, Magnitude and Magnitude Error. We performed analysis on obtained data to establish relation between our output parameters and the selected factors.

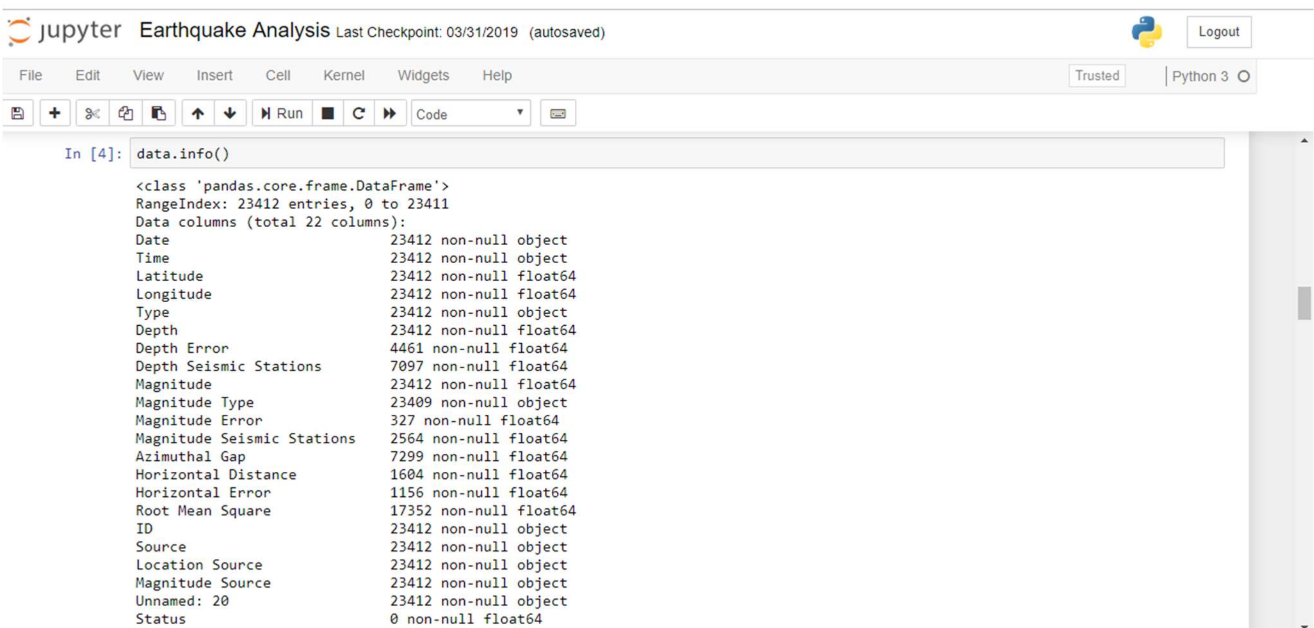
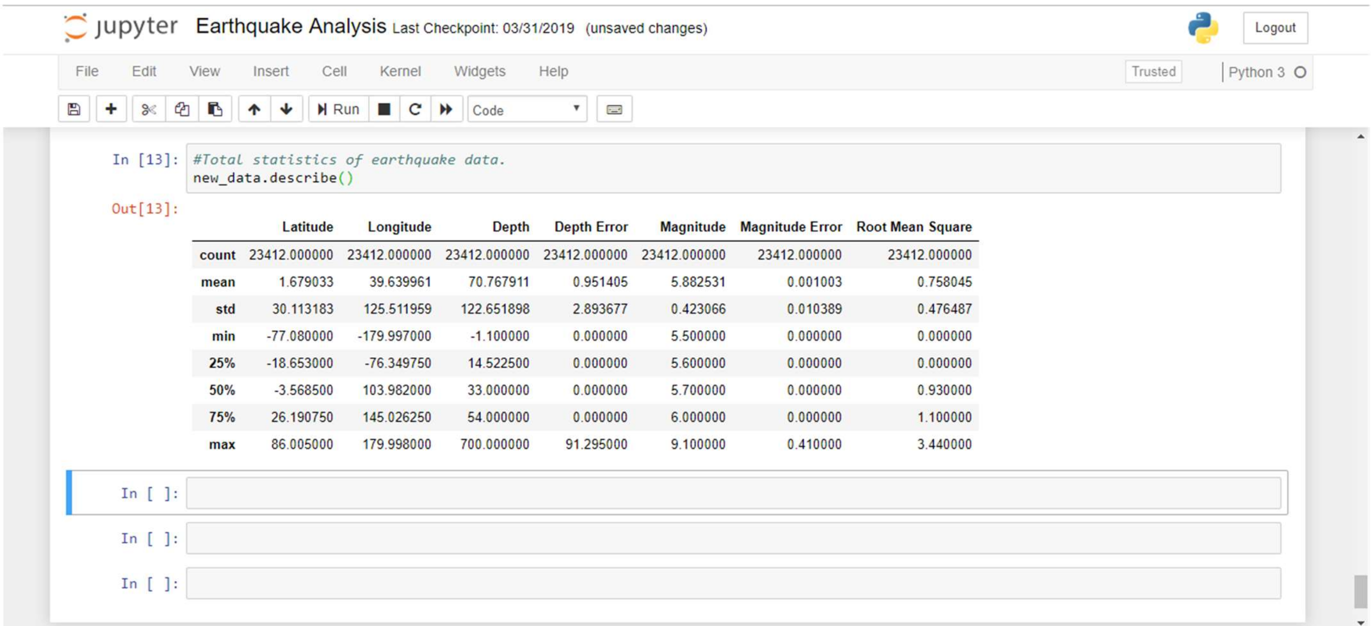


Figure 6.1.1: Information about significant dataset.



In [13]:
```
#Total statistics of earthquake data.
new_data.describe()
```

Out[13]:

	Latitude	Longitude	Depth	Depth Error	Magnitude	Magnitude Error	Root Mean Square
count	23412.000000	23412.000000	23412.000000	23412.000000	23412.000000	23412.000000	23412.000000
mean	1.679033	39.639961	70.767911	0.951405	5.882531	0.001003	0.758045
std	30.113183	125.511959	122.651898	2.893677	0.423066	0.010389	0.476487
min	-77.080000	-179.997000	-1.100000	0.000000	5.500000	0.000000	0.000000
25%	-18.653000	-76.349750	14.522500	0.000000	5.600000	0.000000	0.000000
50%	-3.568500	103.982000	33.000000	0.000000	5.700000	0.000000	0.930000
75%	26.190750	145.026250	54.000000	0.000000	6.000000	0.000000	1.100000
max	86.005000	179.998000	700.000000	91.295000	9.100000	0.410000	3.440000

In [ ]:

In [ ]:

In [ ]:

Figure 6.1.2: Basic statistical measures of the attribute

In this figure 6.1.2 the basic statistical measures are carried out. With the type of the attribute the mathematical measures such as the mean, the mode, the standard deviation, etc. With the following results some desired outputs are to be found.

- There is no missing value.
- The mean for the number of depth and number of depth error is not meaningful.
- According to the max values, there are obviously outliers like for magnitude whose 75% of the data have a median about 7.902, 8.01 appears as the maximum outlier.
- Finally, the minimum value of the magnitude error is 0.001 meaning there is least noisy tuple in the dataset.

- The standard deviation gives also an indication about what should be considered as outliers, but it is not a robust technique since the standard deviation use the mean to be computed.

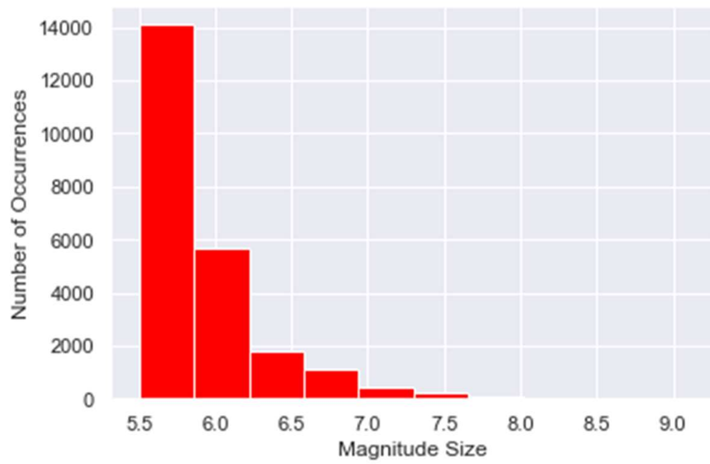


Figure 6.2.3: No of occurrences vs Magnitude Size.

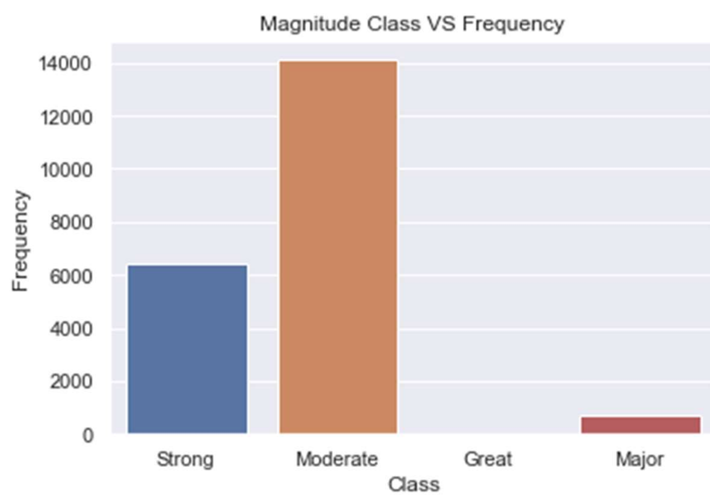


Figure 6.2.4: Magnitude vs Frequency.

From the above figure 6.2.3 the histogram is to plot to understand the relationship between the two parameters No of occurrences and Magnitude Size. From the predicted data number of observations are to be seen. At the magnitude of scale 5.5 the number of occurred earthquakes over worldwide is 14000 and as the Magnitude size increases the number of occurrences are decreasing.

Another figure 6.2.4 shows the relationship between Magnitude Class Vs Frequency in which bar graph represent the type of magnitude over the frequency class. Here at frequency 14000 the Moderate magnitude class were occurred. Major earthquakes were less to be found.

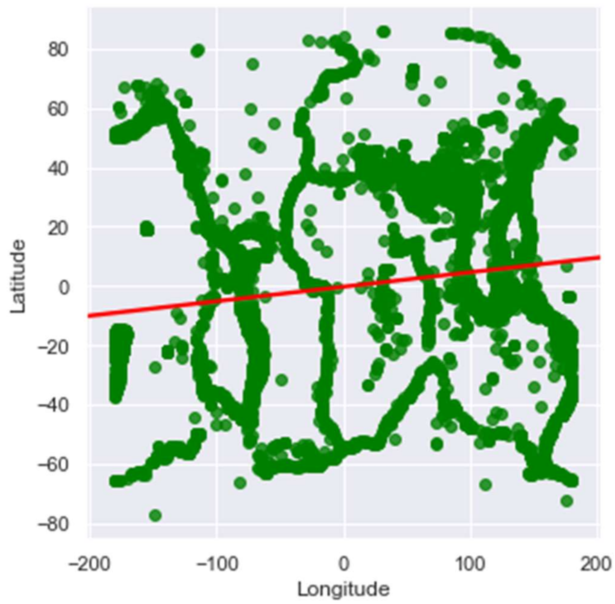


Figure 6.2.5: Latitude vs Longitude

This graph gives the brief information about the geological locations which are plotted in the form of coordinates. For future visualization of the significant data earthquakes on the world map these projections are developed. Depending on the intended use of the map projection, there are certain map features (e.g., direction, area, distance, shape, or other considerations) that are useful to maintain. In this graphs the data points of the earthquakes are to be located so that it becomes easy to visualize the natural disaster earthquake over the world.

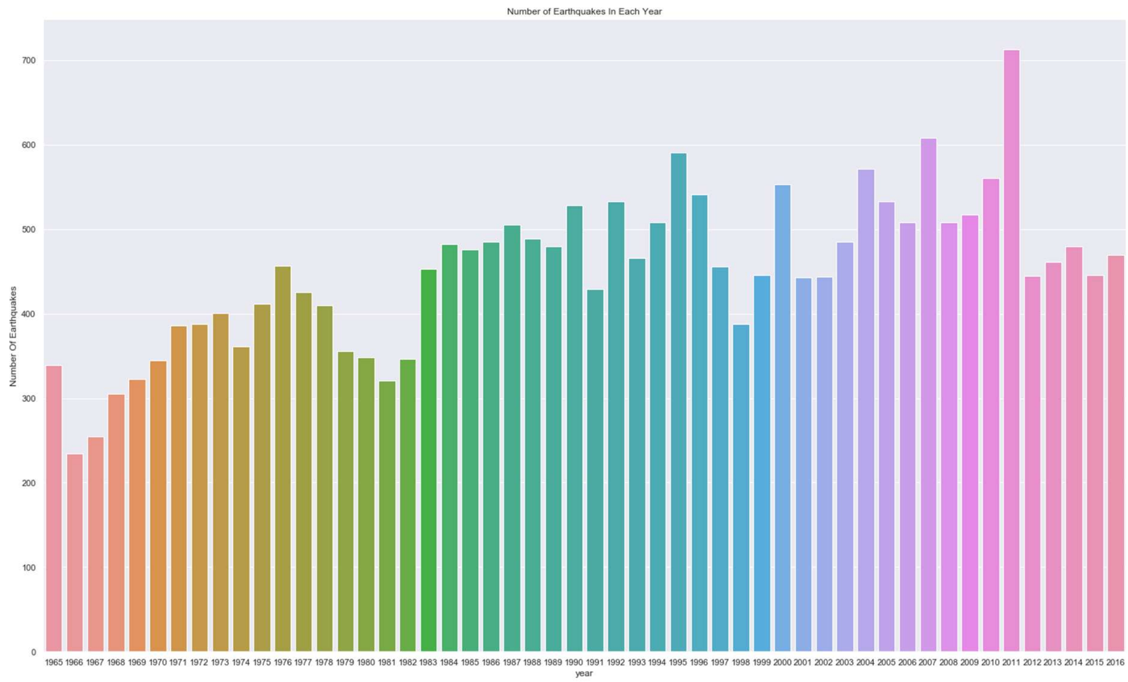


Figure 6.2.6: Earthquakes occurred by year

In this figure 6.2.6 the significant earthquakes occurred over the world is to be illustrated. From year 1965 to 2016 various earthquakes were occurred at different magnitude scale. At year 2011 number of occurrences of the earthquake were more. At specific range of year from 1967 to 2012 the occurrence of this natural disaster were moderate.

6.2 Results

After collection of data of significant earthquake the future magnitude is predicted using neural network. The value is then compared the next day with the actual value. The results of earthquake prediction are illustrated in testing format.

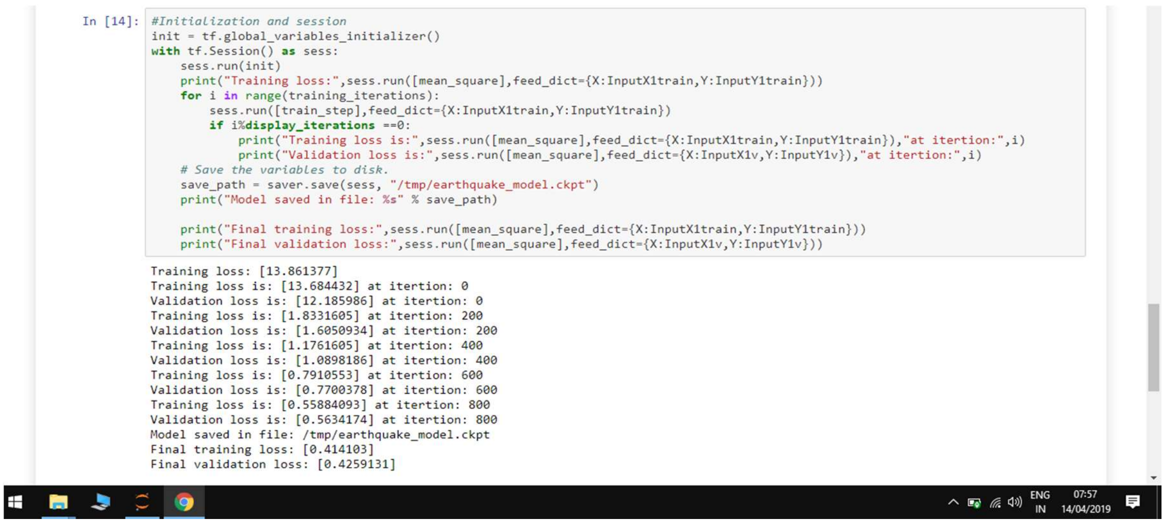


Figure 6.2.1: Training and validating earthquake significant dataset.

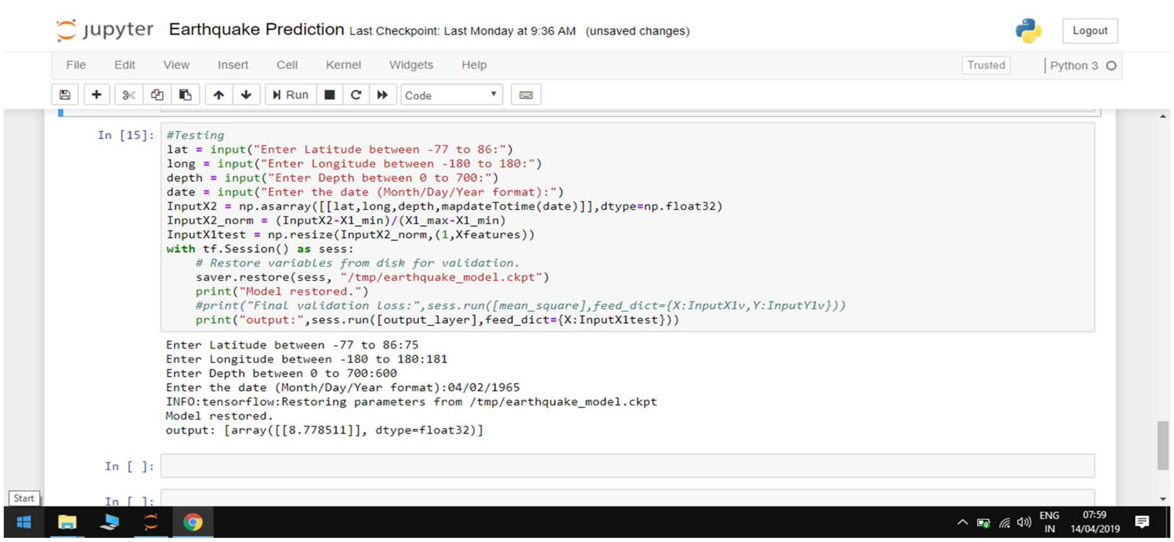


Figure 6.2.2: Prediction of earthquake significant dataset.

Chapter 7

Conclusion

The project has a great potential regarding the earthquake prediction with the help of the help of the attributes considered giving the result in the form of magnitude. To conquer the problem regarding the earthquake taking place in different parts of the world, our project is quite useful in earthquake prediction and avoidance of the loss of life and property. With the help of this neural network model we can not only help the government organisations but also help in the development of the country. Thus, we successfully design a prototype model of earthquake prediction.

Literature Cited

1. Apache Hive https://hive.apache.org/index.html.

2. William H. K. L. Kanamori h. Jennings P, Kisslinger C (2003) International Handbook of Earthquake and Engineering.

3. Alex Holmens, Hadoop in practise manning publication Co.,ISBN.

4. https://kaggle.org/significant earthquake dataset

5. http://www.ibm.com/developerworks/library

Appendix

Technologies Used

1. Python 3.6.14:

Python is a high-level interpreted programming language. It supports object-oriented, imperative, functional and procedural programming. Python interpreters are available for a wide range of operating systems. Python 3.0 was released on 16 October 2000. Rather than having all of its functionality built into its core, Python was designed to be highly extensible. This compact modularity has made it particularly popular as a means of adding programmable interfaces to existing applications. To implement machine-learning algorithms with ease one requires a high-level dynamic programming language. This project uses Python programming language to implement faster RCNN and LSTM algorithms.

2. Tensorflow 1.5.0:

Tensorflow is an open-source software library (framework) which can be for machine-learning applications such as neural networks. The Google Brain team for internal Google use developed TensorFlow. Tensorflow can work using multiple CPUs and multiple GPUs.

3. Anaconda 5.1.0 :

Anaconda is a free and open source distribution of Python programming language for data science and machine learning related applications. Package versions are managed by package management system “conda”. It includes more than 250 popular data science packages available for Windows, Linux and MacOS.

External Libraries

1. **Numpy:-**

   Numpy is a general-purpose array processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. A powerful N-dimensional array object. The core functionality of NumPy is its "ndarray", for *n*-dimensional array, data structure. These arrays are stride views on memory.

2. **Pandas:-**

   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. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.

3. **Matplotlib:-**

   Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.

4. **Seaborn:-**

   Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. It aims to visualization a central part of exploring and understanding data. Its dataset-oriented plotting functions operate on data-frames and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots.

Acknowledgements

We have great pleasure in presenting the report on Earthquake Prediction. We take the opportunity to express my deep regards towards the ones who offered their valuable guidance in the hour of need. Prominent among them are my guide **Mr. Chetan Patil** and our Head of the Department **Dr. G. A. Walikar** . The ideas suggested by them are incorporated in this project topic report. Thank you for supporting me and for making my research possible.

We express our warm thanks to the Principal, **Dr. G. V. Mulgund** for their support and guidance. Furthermore, we would also like to express our special thanks to our friends for the interesting and nice time we had when doing our Graduation together.

Last but not the least we would like to thank our family, especially our parents for always putting my education first and supporting me in all stages of my life.

Digole Suhas Sunil (EU1162095)

Mendapara Harsh Gordhan (EU1162099)

Thakur Manthan Uday (EU1162006)