

Analysis and Prediction of Earthquakes using different Machine Learning techniques

Manaswi Mondol
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
m.mondol@student.utwente.nl

ABSTRACT

A reliable and accurate method for earthquake prediction has the potential to save countless human lives. With that objective in mind, this paper looks into various methods to predict the magnitude and depth of earthquakes. In this paper, real-world earthquake data is analysed to identify patterns and gain insight into this natural calamity. This data is then used to train four machine learning models namely Random forest, linear regression, polynomial regression, and Long Short Term Memory for predicting the magnitude and depth of earthquakes. The performances are compared to find the most effective model. It is very difficult to accurately predict the magnitude of earthquakes however, in this paper it can be seen that polynomial regression shows the best overall results. Also, Random forests are incredibly effective in predicting the depth of an earthquake.

Keywords

Earthquake Prediction, Machine Learning, Regression Analysis, Random Forest, Linear Regression, Polynomial Regression, Long Short Term Memory

1. INTRODUCTION

Earthquakes are devastating natural calamities responsible for widespread death and destruction throughout the world. It is impossible to prevent earthquakes and, their occurrence does not follow any noticeable pattern. A reasonably accurate and timely prediction of an earthquake can potentially save thousands of lives and also prevent damage to property. In the last two decades, countless studies have been conducted on the topic of earthquake prediction. Predicting an earthquake involves stating the exact time, magnitude and location of an upcoming earthquake. This prediction can be classified into short-term, intermediate and long-term prediction based on the duration of the time scale. This problem is nigh impossible to solve and as a result in 1997, Geller et al. [5] concluded that predicting earthquakes is impossible. Over time research has shown that it is not completely impossible to predict earthquakes. In this paper, some of the studies that have been successful in developing a reasonably accurate prediction model have been discussed [2, 8, 10, 3,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

35th Twente Student Conference on IT July 2nd, 2021, Enschede, The Netherlands.

Copyright 2021, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

9]. However, in spite of such great effort by scientists throughout the world a valid and reliable method to make perfectly accurate prediction has not been found yet. In this paper, the general problem of earthquake prediction is narrowed down to predicting the magnitude and depth of an earthquake.

In this paper the following research question will be answered

- *How can machine learning techniques be used to predict the approximate magnitude and depth of earthquakes by analyzing earthquake data?*

In order to answer this question analysis of the data set will be done followed by training different machine learning models and finally comparing their performances. Consequently, the following sub research questions will be answered.

- *Which parameters of earthquake data are relevant for this research?*
- *What insights can be obtained from analysing the data ?*
- *Which machine learning techniques are most effective and produces the best results?*

This research attempts to analyse earthquake data and predict the magnitude and depth of earthquakes. The analysis of earthquake data performed here helps to identify the important attributes of the data which can be used for prediction. The analysis also provides some interesting insights like the existence of outliers, the most frequent earthquake magnitude, and the distribution of earthquakes on the world map based on the location and magnitude. Four machine learning models were trained using the data set namely: random forest, linear regression, polynomial regression, and Long Short Term Memory. The results from the different models are compared to identify the best performing model.

In section 2, the different machine learning models used in this research are briefly explained. Prior research conducted on this topic is also explored in this section. In section 3, exploratory data analysis and data cleaning is performed on the data set to make sure that relevant data is used to train the machine learning models. In section 4, the machine learning models are trained using the data and cross validation is performed to find the best hyperparameters. In section 5, the results from the different models are evaluated and compared to find the best-performing machine learning model.

2. BACKGROUND

2.1 Regression Analysis

The magnitude and depth of an earthquake being numerical data, regression analysis will be used for the purpose of prediction. Regression analysis is a statistical method used for finding relationships between independent and dependent variables. Regression analysis can be used for prediction and forecasting. In this paper, regression analysis will be used to predict the magnitude and depth of earthquakes. Linear regression is the most widely used regression algorithm where a linear model is fitted to explain the relationship between the independent and dependent variables[1]. When linear regression has to be performed for two or more dependent variables, multiple linear regression is used [1]. Multiple linear regression will be used in this research to predict both magnitude and depth.

Sometimes, a straight line is unable to properly explain the relationship between dependent and independent variables. In cases like that fitting a polynomial of degree n to the data can perform significantly better and help model non-linear relationships. This is referred to as polynomial regression and can be viewed as an extension of the simple linear regression model [1]. Other Non-linear regression models like random forests and neural networks will also be explored.

2.2 Random Forest

Decision trees are non-parametric supervised machine learning techniques used for classification and regression. Random forest is a supervised machine learning algorithm based on decision trees but has significantly better performance. Random forest builds a large number of decision trees and merges them through bagging to produce more accurate and stable predictions[4]. Random forests are very versatile as they are effective for both classification and regression problems. A common problem for machine learning algorithms is overfitting of data however, this problem can be easily solved in a random forest model if there are sufficient number of decision trees[4]. Because of these reasons a random forest model will be used in this research.

2.3 LSTM

Neural networks are modelled after the human brain and contain sensory units called neurons which are connected to each other through weights[15]. Neural networks are very effective in modelling complex non-linear relationships. Recurrent neural networks are a type of neural network that is capable of tracking information from previous states or sequences[15]. RNNs are very effective when dealing with sequential data. LSTM or Long Short-Term Memory is a recurrent neural network architecture that solves the problem of vanishing gradient[7]. An LSTM network has four neural networks, three of which are control gates namely - the input, output, and the forget gate and the fourth neural network is used to estimate the memory cell parameter[15]. The introduction of these gates helps to improve the learning capabilities and the memory capacity of an LSTM network. This makes LSTM networks very useful when dealing with large sequential data.

2.4 Literature Review

The occurrence of earthquakes is a highly random phenomenon and so the existence of a model that can ac-

curately predict the time, location and magnitude of an earthquake is not known. Several studies have been conducted throughout the years by researchers in this field. Researchers have tried to approach this problem from different perspectives to find a potential solution. In this section, some of these studies which are relevant and helpful for this research will be discussed.

Since this research involves predicting the magnitude and depth of earthquakes, the paper written by Mallouhy et al. [10] serves as a good starting point. In this paper, eight different machine learning algorithms are implemented to predict if an earthquake event can be classified as major or minor. This paper helps to understand different machine learning algorithms when these are applied to the context of earthquake data. Random forest and K nearest neighbour algorithm showed the best results when compared to other models[10].

There are two general approaches to predict earthquakes namely precursor based and trend based[11]. Precursors are phenomena or signals that precede an impending earthquake for e.g; radon gas emissions and unusual animal behaviour. A great example of precursor based prediction can be found in the paper by Kuyuk et al. [8]. In this paper, they explain how they were able to train a Long Short Term Memory (LSTM) to analyze information collected from Earthquake Early Warning systems to detect earthquakes before the seismic waves reach the centre of a city or a densely populated area[8]. This can potentially save many lives as being aware of an earthquake even a bit earlier can drastically improve evacuation measures [8]. Trend based methods deal with identifying patterns in real-world data e.g; seismicity, prior earthquakes etc. to predict earthquakes. This research also involves trend-based prediction hence finding similar studies will prove helpful.

In their paper, Asim et al.[2] predicted the magnitude of earthquakes in the Hindukush region using historic seismic data. They applied four machine learning algorithms on a data set namely - pattern recognition neural network, recurrent neural network(LSTM), random forest with 50 trees and linear programming boost ensemble [2]. Since the study focused on predicting earthquakes in the same location, the pattern recognition neural network showed the best results with an accuracy of 65%[2].

Li et al.[9] explore the possibility of predicting aftershocks with a magnitude greater than 4.0 in their research by proposing a new model called PR-KNN. PR-KNN is a combination of the polynomial regression method and the K nearest neighbours(KNN) algorithm. This proposed method was applied to experimental data from the Wenchuan website. PR-KNN achieved noticeably better results than the traditional KNN and Distance-weighted KNN algorithms.

Lastly in the study conducted by Bhandarkar et al., they predicted earthquake trends using Long Short Term Memory (LSTM) and Feed Forward Neural Network (FFNN) and compare the results[3]. This study is very similar to this one and was able to provide valuable information regarding the performance of LSTM networks in this context. The results from this paper show that an LSTM network is 59% more effective than an FFNN [3].

3. METHODOLOGY

The United States Geological Survey (USGS) provides real-world earthquake data on past earthquakes. The USGS classifies “significant events” based on a combination of the magnitude of an earthquake, the number of ‘Did You Feel It’ responses, and PAGER alert level[14]. The decision to work with only significant earthquakes was motivated by two reasons namely, the time constraint and the reduction in the volume of the data used to train the machine learning models. The data set used for this research was obtained from www.kaggle.com/usgs/earthquake-database which contains earthquake data collected from the USGS website by Kaggle. This data set includes a record of the date, time, location, depth, magnitude, and source of every earthquake with a reported magnitude of 5.5 or higher from 1965 until 2016.

3.1 Data Cleaning

The data set contains the following attributes describing earthquakes - ‘Date’, ‘Time’, ‘Latitude’, ‘Longitude’, ‘Type’, ‘Depth’, ‘Depth Error’, ‘Depth Seismic Stations’, ‘Magnitude’, ‘Magnitude Type’, ‘Magnitude Error’, ‘Magnitude Seismic Stations’, ‘Azimuthal Gap’, ‘Horizontal Distance’, ‘Horizontal Error’, ‘Root Mean Square’, ‘ID’, ‘Source’, ‘Location Source’, ‘Magnitude Source’, and ‘Status’.

Out of these attributes ‘Date’, ‘Time’, ‘Latitude’, ‘Longitude’, ‘Type’, ‘Depth’, ‘Magnitude’, ‘Magnitude Type’, ‘ID’, ‘Source’, ‘Location Source’, ‘Magnitude Source’, and ‘Status’ have no null values. The attributes ‘Magnitude Type’, ‘ID’, ‘Source’, ‘Location Source’, ‘Magnitude Source’, and ‘Status’ are categorical variables and thus have no significance in regression analysis. Hence, the attributes ‘Date’, ‘Time’, ‘Latitude’, ‘Longitude’, ‘Magnitude’, and ‘Depth’ will be used for regression analysis in this research.

The built-in functions in the pandas library for python were used for identifying the number of null and non-null values for each attribute. A detailed overview of this can be found in table 1

3.2 Exploratory Data Analysis

A total of 23412 earthquakes are recorded in this data set. By analysing the magnitude types of different earthquakes it can be seen that 99.5% of all the earthquakes fall under six categories as observed in table 2 [13]. The different Magnitude types can be explained as follows:

- MW (Moment Magnitude) - Derived from seismic moment (most common and general type).
- MWC (Moment Centroid) - Derived from a centroid moment tensor inversion of long-period surface waves.
- MWB (Moment Body Wave) - Derived from a centroid moment tensor inversion of body waves.
- MB (Moment Short Body Wave) - Derived from a centroid moment tensor inversion of short-period body waves.
- MWW (Moment W-phase) - Derived from a centroid moment tensor inversion of the W-phase.
- MS (Moment Surface Wave) - Derived from a centroid moment tensor inversion of surface waves.

It makes sense to study the basic statistics regarding the attributes magnitude and depth. The relevant values can

be found in the table 3. First, the outliers in the data set will be detected based on magnitude and depth values. Outlier detection can be done using the interquartile range(IQR) and the 3-sigma rule.

IQR is equal to the difference between 75th(Q3) and 25th(Q1) percentiles. Outliers are defined as observations that fall below $Q1 - 1.5 * IQR$ or above $Q3 + 1.5 * IQR$.

Earthquakes having magnitude outside this range (4.99, 6.60) are considered outliers. This can be seen in the figure 1. The earthquake with highest magnitude (9.1 on the Richter scale) in this data set was recorded on December 26, 2004 and is known as the Sumatra-Andaman earthquake [13].

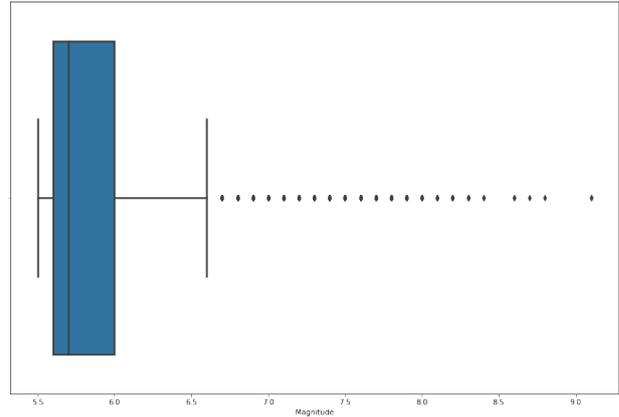


Figure 1. Boxplot: Earthquake magnitude

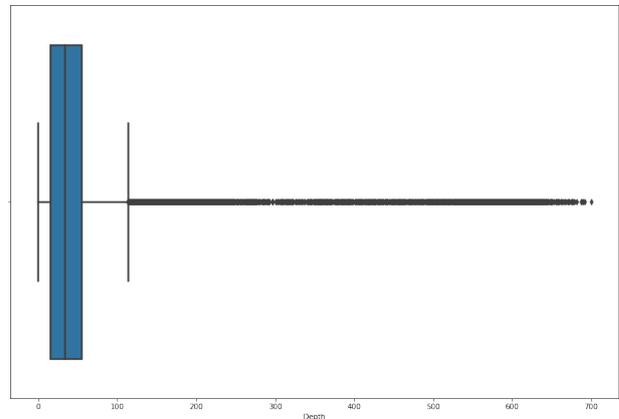


Figure 2. Boxplot: Earthquake depth

Earthquakes detected having depth outside of this range (−44.69, 113.21) are considered outliers. This can be seen in figure 2.

The 3-sigma rule or the empirical rule, states that 68%, 95%, and 99.7% of the values in a data set lie within one, two, and three standard deviations of the mean, respectively. Thus the absolute value of the z-score of the magnitude and depth of an earthquake should be less than 3. Using this 3-sigma it can be seen that 1050 and 447 earthquakes in this data set are classified as outliers based on magnitude and depth values respectively.

In figures 3 and 4, the number of occurrences of earthquakes with varying magnitudes are plotted. From figure 4, it is clearly visible that most of the earthquakes

Table 1. Data set null values

Values	Date	Time	Latitude	Longitude	Depth	Depth Error	Depth Seismic Station	Mag	Mag Type	Mag Error	Mag Seismic Station	Azimuthal Gap	Hor. Dist	Hor. Error	Root Mean Square	ID	Source	Location Source	Magnitude Source	Status
Non-Null	23412	23412	23412	23412	23412	4461	7097	23412	23409	327	2564	7299	1604	1156	17352	23412	23412	23412	23412	23412
Null	0	0	0	0	18951	16315	0	3	23805	20848	16113	21808	22256	6060	0	0	0	0	0	0

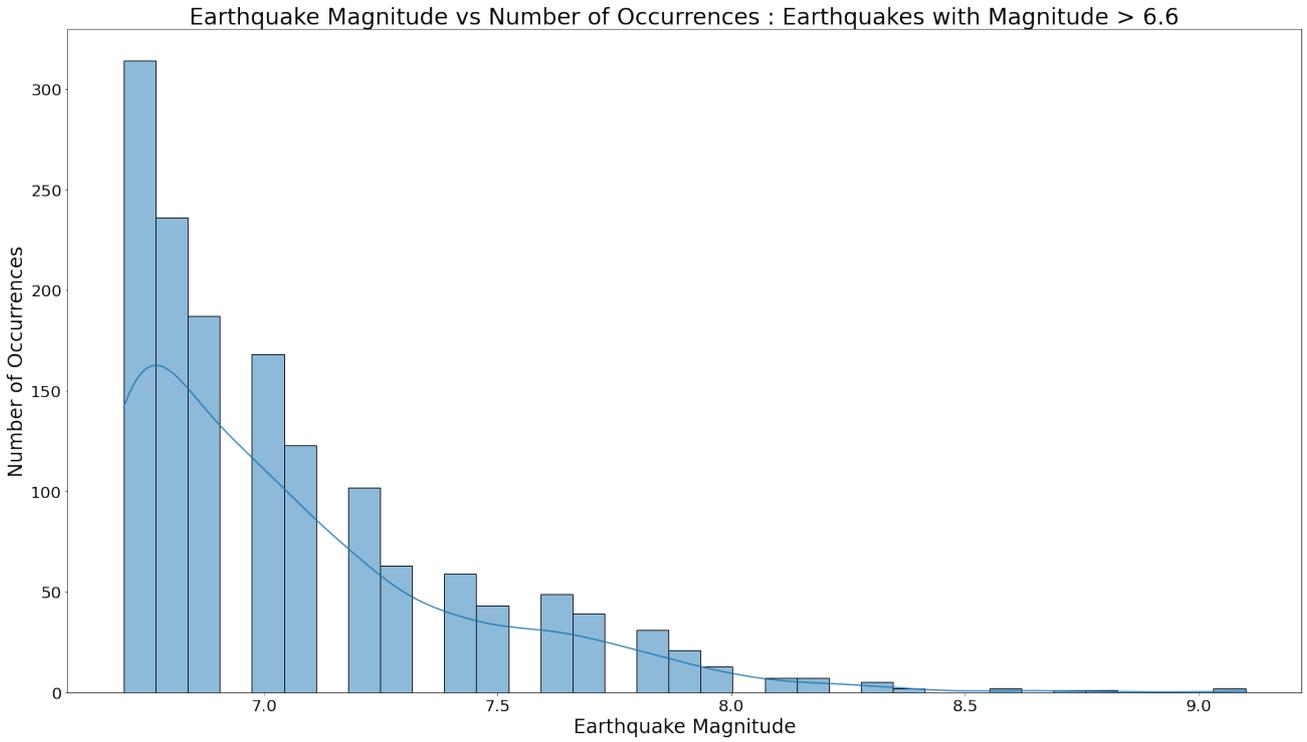


Figure 3. Earthquake Magnitude vs Number of Occurrences

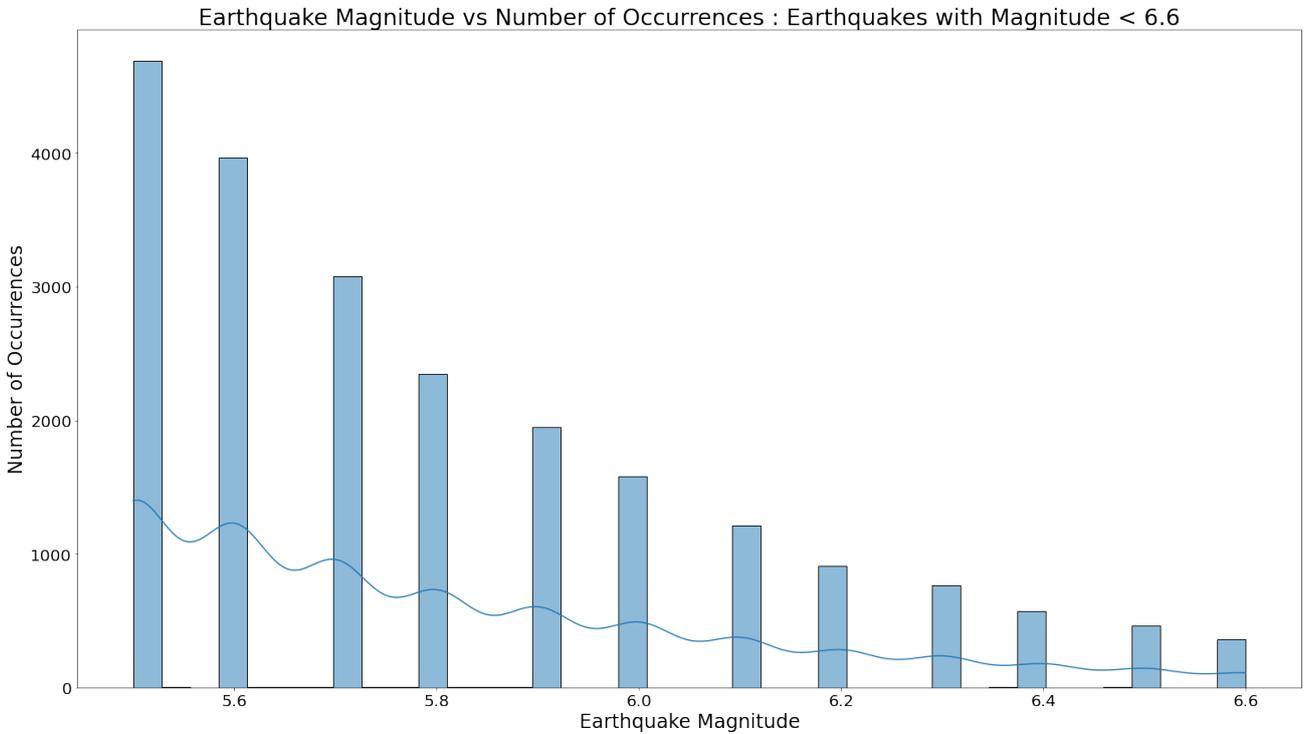


Figure 4. Earthquake Magnitude vs Number of Occurrences

have a magnitude between 5.5 and 6.6. In figure 3, it can be seen that relatively few earthquakes have magnitude

higher than 6.6 which is consistent with the boxplot in figure 1.

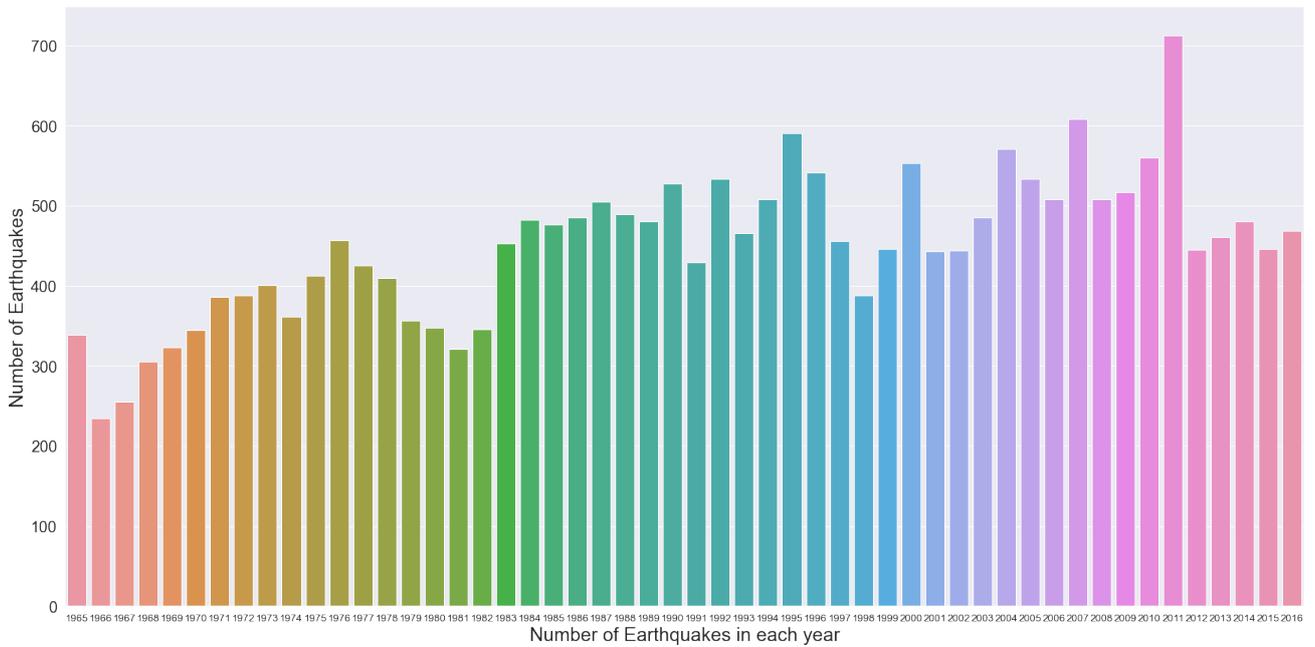


Figure 5. Number of Earthquakes in each year

Table 2. Different Magnitude Types

Magnitude Type	MW	MWC	MB	MWB	MWW	MS	Others
Percentage (%)	33	24.2	16	10.5	8.5	7.3	0.5

Table 3. Magnitude and Depth statistics

	count	mean	std	min	25%	50%	75%	max
Magnitude	23412	5.88	0.42	5.5	5.6	5.7	6.6	9.1
Depth	23412	70.77	122.65	-1.1	14.5	33	54	700

3.3 Understanding the Data

The number of earthquakes that occurred each year from 1965 to 2016 can provide some important insight into the data. In figure 5, a count plot shows the number of earthquakes in each year. From this count plot it can be seen that in the last 50 years the highest number of earthquakes were recorded in 2011 with 713 earthquakes followed by 2007 when a total of 608 earthquakes were recorded. In 1996, the lowest number of earthquakes were detected with a total of 234.

It is necessary to know if the two attributes magnitude and depth are correlated and the possibility to identify the underlying relationship. The correlation coefficient between the magnitude and depth of the earthquakes recorded in the data set yields a value of 0.0234. If the value of the correlation coefficient approaches 0 it can be interpreted as the two attributes having no correlation at all. It can be concluded that in the context of this data set magnitude and depth are not correlated. An earthquake can be detected on the earth's surface anywhere between 0 – 700 km and is categorized as shallow, intermediate or deep based on the depth. These earthquakes can be detected on land and also on the ocean floor. However, earthquakes of similar magnitude can have varying depths if the detected location is below the ocean surface as compared to on land. Out of the 23412 earthquakes in the data set, 21937 earth-

quakes have a magnitude lower than 6.6 and the remaining 1475 earthquakes have magnitudes higher than 6.6. This information can be visually represented by plotting the latitude and longitude of different earthquakes on a world map. This provides a better context to the data and helps understand how different earthquakes are distributed throughout the Earth based on their source.

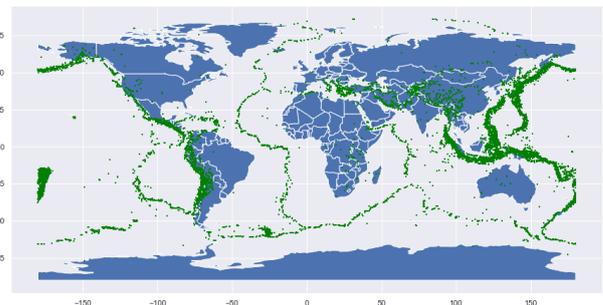


Figure 6. Earthquakes with magnitude < 6.6

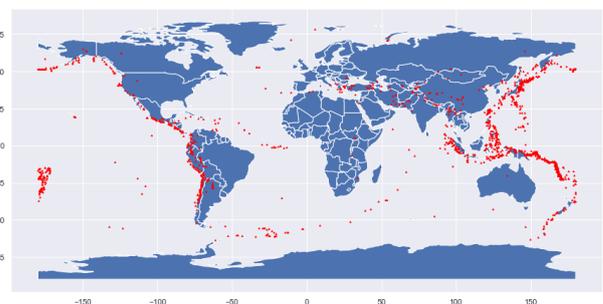


Figure 7. Earthquakes with magnitude > 6.6

In figure 6, the earthquakes with magnitude less than 6.6 are shown and they are represented by a green marker. And in figure 7, the earthquakes with magnitude greater

than 6.6 are shown and they are represented by a red marker.

4. EXPERIMENT

The following four models were trained using the data -

1. Random forest
2. Linear regression
3. Polynomial regression
4. Long Short Term Memory

The date and time of occurrence of an earthquake does not follow a pattern and the interval between two subsequent earthquakes is never same. Hence the data cannot be considered to a time-series. For the models the input will be Latitude and Longitude of the earthquake and output will be the magnitude and depth. The data is split between training and testing sets with the training data set containing 80% and the testing data set containing 20% of the data.

4.1 Random Forest

The built-in random forest regression function from the library scikit-learn is used. The model is trained at first with 10 decision trees as this is the default value.

In supervised machine learning when a model works very well on the training data but is unable to perform similarly on the testing data the situation called overfitting. This problem can be solved by finding the optimum values of the hyperparameters of a model. A hyperparameter is a parameter whose value is used to control the learning process of a model. In case of random forest this hyperparameter is the number of decision trees generated during the learning process. The built-in 'GridSearchCV' function in scikit-learn will be used for the process of tuning hyperparameters.

Since the default value number of trees is 10, the model is further trained with the number of decision trees generated ranging from 20 to 200 with an interval of 10. After running this process it is observed that the model shows the best performance when 120 decision trees are used.

Random Forest uses decision trees for regression analysis and hence does not require feature scaling. The algorithm partitions the data set so even if feature scaling like min-max normalization is applied the results would remain unchanged.

4.2 Linear Regression and Polynomial regression

A linear regression model is trained using the data and the built-in functions of scikit-learn are used. Linear regression and polynomial regression are sensitive to high values. Due to this reason, feature scaling is important as the attributes in the data set are measured in different units which vary across a wide range and this might end up creating a bias. Min-max normalization is used on the data set to make sure that all values are within the range (0,1).

For polynomial regression, the relationship between dependent and independent variables is explained using a polynomial of degree n . The degree parameter is varied from 2 to 20 and some interesting results can be seen. A polynomial of degree 16 shows the best results.

4.3 Long Short Term Memory

The keras library is used to implement the LSTM model in this research. Like in the previous models we will try to predict the magnitude and depth of an earthquake using regression analysis. Min-max normalization is used on the data set. The data set is split into training and testing sets containing 80% and 20% of the data respectively with 10% of the dataset used for validation.. For an LSTM network we have to set the lookback value which is the number of previous inputs the model will take into account when predicting the next value. The model is tested with lookback values 50 and 100 respectively.

The Stacked LSTM is an extension to a standard LSTM model which can have multiple hidden LSTM layers[6]. The addition of hidden layers makes the model deeper and helps identify the complex relationship between multiple variables in a data set[6]. The network used in this research has a visible layer, 2 hidden layers with 100 neurons and with a sigmoid activation function, and an output layer that predicts 4 values. The linear activation function is used in the output layer as this does not change the weighted sum of the input and returns the direct value.

Dropout is a regularization technique for neural network models where randomly selected neurons are ignored during training[12]. This makes the model less sensitive to neuron weights and also prevents overfitting [12]. Dropout with a 20% (i.e. 20% of the neurons are discarded each weight cycle)is implemented in this model using the built-in dropout function in keras.

5. RESULTS

5.1 Performance Metrics

R^2 score, explained variance score ($ExVar$) and mean squared error(MSE) will be used to evaluate the performance of the models implemented in this research.

R^2 score or the coefficient of determination determines the effectiveness of the regression model and is defined by :

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ and $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \epsilon_i^2$ \hat{y} is the predicted value and y is the corresponding true value.

Explained variance score is defined as :

$$Ex_var(y, \hat{y}) = 1 - \frac{Var\{y - \hat{y}\}}{Var\{y\}}$$

where \hat{y} is the estimated target output, y the corresponding true target output, and Var is Variance, the square of the standard deviation.

Mean squared error measures the average of the square of errors and is defined as :

$$MSE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2$$

where \hat{y}_i is predicted output and y_i is true output of the i th sample.

5.2 Discussion

Table 4 shows the evaluation metrics for all the models that have been tested and their relative effectiveness when trying to predict the magnitude and depth of earthquakes.

R^2 score and explained variance score determines the

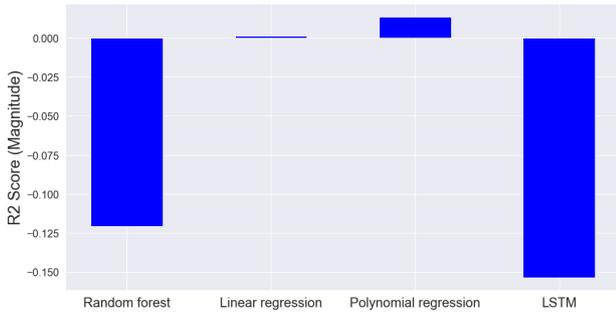


Figure 8. R^2 score : Magnitude

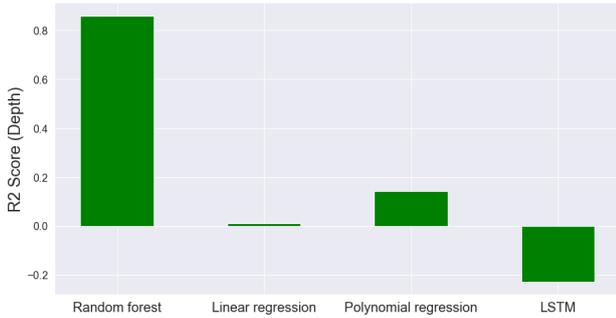


Figure 9. R^2 score : Depth

regression score of a model and can achieve a best possible score of 1.0. A score of 0.0 indicates that the model always predicts the expected value of the output, disregarding the input features. A negative value implies that a straight line fits the data better than the tested model. This can be seen from the scores in table 4. Both R^2 score and explained variance score attain a negative value while predicting magnitude in case of random forest and LSTM. This can be seen in figure 8 where it also shows that linear regression performs better in comparison.

In figure 12, the relative performances of the polynomial regression is compared. The performance gradually increases from degree 2 to degree 12 and then falls abruptly from there. A polynomial of degree 16 shows the best performance. Coincidentally, polynomial regression with a polynomial of degree 16 is also the best performing model showing comparatively better results than others. For predicting the magnitude of an earthquake polynomial regression yields the best results followed by linear regression, random forest, and LSTM. This can be seen from figures 8 and 10, where the R^2 scores and mean squared error for magnitude prediction are shown.

However, an interesting result is obtained from the random forest model. The random forest model is able to predict the depth of an earthquake incredibly well. Both R^2 score and explained variance scores in this case are 0.8574 which is very close to the perfect value of 1.0 as seen in figure 11. The mean squared error for depth is also the lowest for random forest when compared to the other models as seen in figure 11.

6. CONCLUSION

The performance from four different machine learning models used in this research further supports the argument that it is very difficult to accurately predict earthquakes. Analysis of the data set provided great insight and helped

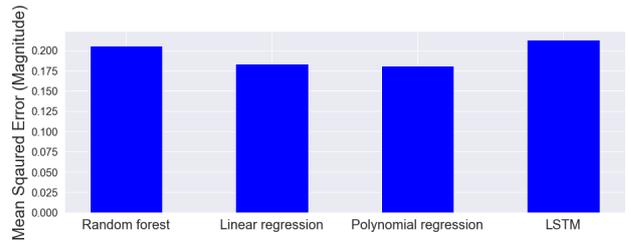


Figure 10. Mean squared error : Magnitude

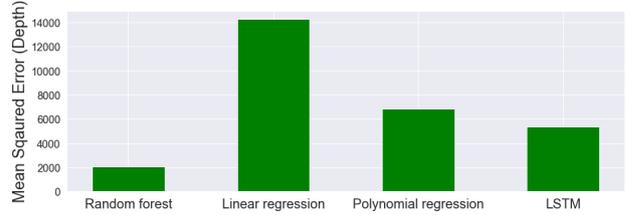


Figure 11. Mean squared error : Depth

identify patterns and key attributes that would be relevant for this research. Latitude and longitude of detected earthquakes can be used as dependent variables to predict the magnitude and depth of future earthquakes. This can be helpful for determining the magnitude and depth of earthquakes in seismic zones where the frequency of recurrent earthquakes is high.

Furthermore, this paper presents three important findings

1. Polynomial regression (with a polynomial of degree 16) is the best method for predicting the magnitude of an earthquake.
2. Random forest can be extremely effective in predicting the depth of earthquakes
3. Polynomial regression is the overall best performing model.

For further improvement seismic data collected from seismographs positioned all around the world can be analysed and used to improve the different models implemented here. Also, given the nature of polynomial regression and the size of the data set used, there is a high probability of overfitting in the polynomial regression model used in this research. This can be another topic for further improvement.

7. REFERENCES

- [1] ANDREWS, D. F. A robust method for multiple linear regression. *Technometrics* 16, 4 (1974), 523–531.
- [2] ASIM, K., MARTÍNEZ-ÁLVAREZ, F., BASIT, A., AND IQBAL, T. Earthquake magnitude prediction in hindukush region using machine learning techniques. *Natural Hazards* 85 (01 2017), 471–486.
- [3] BHANDARKAR, T., K, V., SATISH, N., SRIDHAR, S., SIVAKUMAR, R., AND GHOSH, S. Earthquake trend prediction using long short-term memory rnn. *International Journal of Electrical and Computer Engineering (IJECE)* 9 (04 2019), 1304.
- [4] BREIMAN, L. Random forests. *Mach. Learn.* 45, 1 (Oct. 2001), 5–32.

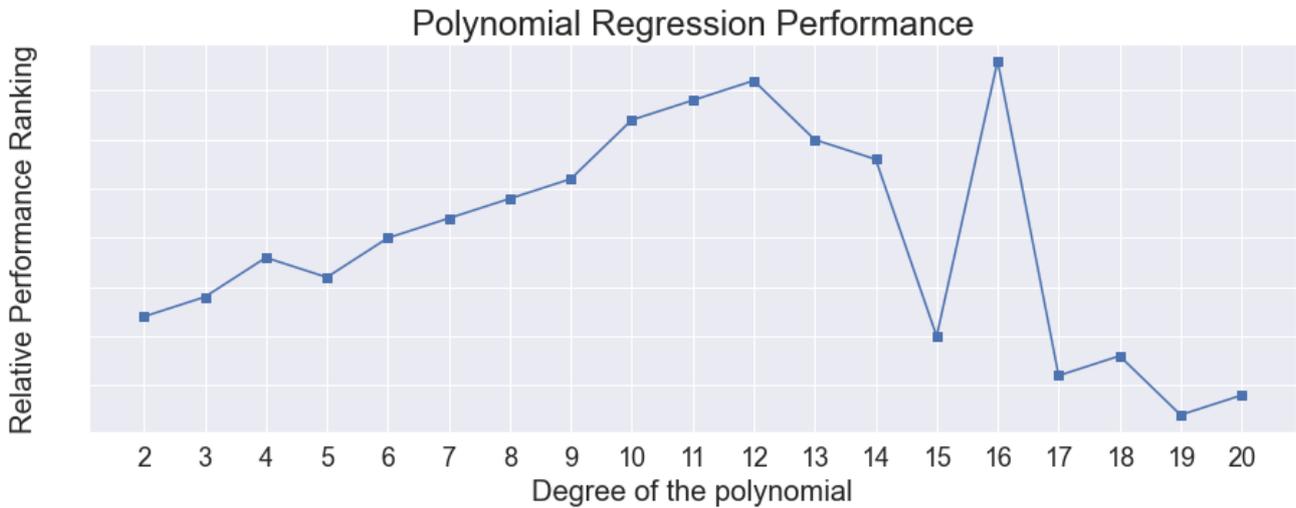


Figure 12. Polynomial Regression Performance Comparison

Table 4. Performance : Polynomial regression, Linear regression, Random forest, and LSTM

Model/Metrics	R^2 score (Mag)	R^2 score (Depth)	ExVar (Mag)	ExVar (Depth)	MSE (Mag)	MSE (Depth)
Poly Reg	0.0132	0.1416	0.0132	0.1420	0.1809	6862
Linear Reg	0.0013	0.0102	0.0013	0.0103	0.1831	14308
Random Forest	-0.1207	0.8574	-0.1207	0.8574	0.20550	2061
LSTM	-0.1553	-0.2275	-0.0223	-0.0006	0.2131	5350

- [5] GELLER, R., JACKSON, D., KAGAN, Y., AND MULARGIA, F. Earthquakes cannot be predicted. *Science* 275 (1997), 1616 – 1616.
- [6] GRAVES, A., MOHAMED, A.-R., AND HINTON, G. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (2013), pp. 6645–6649.
- [7] HOCHREITER, S., AND SCHMIDHUBER, J. Long short-term memory. *Neural computation* 9 (12 1997), 1735–80.
- [8] KUYUK, H. S., AND SUSUMU, O. Real-time classification of earthquake using deep learning. *Procedia Computer Science* 140 (2018), 298–305.
- [9] LI, A., AND KANG, L. Knn-based modeling and its application in aftershock prediction. In *Proceedings of the 2009 International Asia Symposium on Intelligent Interaction and Affective Computing* (USA, 2009), ASIA '09, IEEE Computer Society, p. 83–86.
- [10] MALLOUHY, R., ABOU JAOUDE, C., GUYEUX, C., AND MAKHOUL, A. Major earthquake event prediction using various machine learning algorithms. In *International Conference on Information and Communication Technologies for Disaster Management* (Paris, France, Dec. 2019).
- [11] SHEARER, P. M. *Introduction to Seismology*, 2 ed. Cambridge University Press, 2009.
- [12] SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I., AND SALAKHUTDINOV, R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15 (06 2014), 1929–1958.
- [13] USGS. Earthquake hazards. www.usgs.gov/natural-hazards/earthquake-hazards/earthquakes. "Accessed: 2021-17-06".
- [14] USGS. Significant earthquakes - 2021. <https://earthquake.usgs.gov/earthquakes/browse/significant.php>. "Accessed: 2021-17-06".
- [15] ZHANG, A., LIPTON, Z. C., LI, M., AND SMOLA, A. J. *Dive into Deep Learning*. 2019. <http://www.d2l.ai>.

APPENDIX

A. PERFORMANCE METRICS

Table 5. Performance Ranking: Random forest

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
No of Trees	120	200	190	180	170	160	150	140	110	30	20	10	60	130	50	70	80	40	100	90

Table 6. LSTM: 2 hidden layers with 50 neurons (lookback = 50)

LSTM	R^S_{core} (Mag)	R^S_{core} (Depth)	Ex_Var (Mag)	Ex_Var (Depth)	MSE (Mag)	MSE (Depth)
50-50	-0.3830	-0.1911	-0.3324	-0.0008	0.3152	7476

B. FIGURES

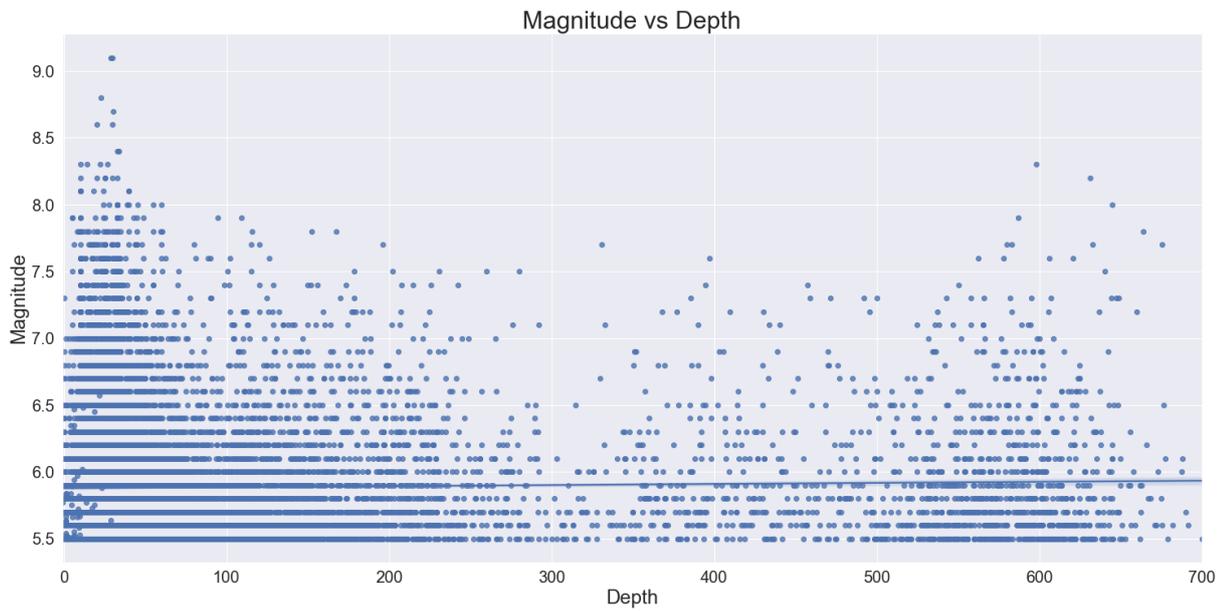


Figure 13. Scatter plot : Magnitude vs Depth