PalArch's Journal of Archaeology of Egypt / Egyptology

Perception And Administering Regression Analysis On Earthquake Data For Finding An Appropriate Correlation: Case Study On Nepal Earthquake

Ms. Sumita mukherjee[1],Dr. Prinima gupta[2],Prof. Felix musau[3]

[1]research scholar, manav rachna university, faridabad, haryana, india

Director, university advancement & internationalization, riara university, nairobi, kenya.

[2]professor, manav rachna university, faridabad, haryana, india.

[3]dean, computing science, riara university, nairobi, kenya.

Email: [1]smukherjee@riarauniversity.ac.ke, [2]prinima@mru.edu.in
[3]fmusau@riarauniversity.ac.ke

Ms. Sumita Mukherjee, Dr. Prinima Gupta, Prof. Felix Musau: Perception and Administering Regression analysis on Earthquake Data for finding an appropriate Correlation: Case study on Nepal Earthquake -- Palarch's Journal Of Archaeology Of Egypt/Egyptology 17(9). ISSN 1567-214x

Keywords: Linear Regression, Earthquake Prediction, Multiple Regression, magnitude, earth's crust, seismic waves, Google Earth, Big Data

ABSTRACT

A popular statistical method, Regression Analysis is built between a dependent variable and more than one independent variable/s to generate the relationship and correlation which can be used for varied predictions. For estimating the same, regression analysis is applied statistically, mathematically and justifies to conclude a logical result calculating Residual Value (RV) and R Square in regression for earthquake prediction. The primary focus of this paper is to produce better prediction on earthquake in terms of resilience, versatility, accuracy, authenticity so that emphasis is applied on nonlinear, linear and multilinear regression model having free hand curve, least square method, standard deviation (SD) using arithmetic mean and assumed mean and statistical analysis. Keeping this result in mind, multilinear regression model is achieved by using historical data of Nepal statistically for five years with parameters like the magnitude, the location, the date, the time, the depth etc. and established a relationship like the dependency of depth and magnitude on earth's crust and temperature on probability of occurrences of earthquake through ANOVA table. This study also recommends that that seismic waves of the earth's crust dependency on three of the parameters are analyzed are very significant predictors for the occurrences of earthquake. We could also find significant correlations between the analyzed indicators..

1. Introduction

The effect of a natural hazards on lives are undescribed as the natural activities cause them imposes a natural disaster. Once earthquake is discussed it is an unexpected motion of the crust of earth which begins below or at the earth's surface. The first covering has various sections named as plates appears as outer coverings and solid in nature. The focus and origin of the seismic movements takes place at mantle also recognized as crust of earth. This is quite extensive placed vertically at the crust of earth popularly known as epicenter.

Volcano eruptions, Flashflood, Wild fire, Earthquakes, Tsunamis, Earthquakes, Artic warming, Landslides, are composite features and components which head for loss of precious human lives, monetary loss, environment challenges, mental imbalance. The call for the day is to save and preserve such environmental and geological disasters. The research studies also seek the support of meteorological, seismological, environmental departments to provide the authentic data so that the available data can serve as a flexible parameter to continue research studies. The data available is a big storage pool termed as big data. The required, accurate and informative data are extracted from big data and the appropriate data mining algorithms are chosen to analyze. These analyzes lead to prediction. There are different and varied approaches survive which can be traversed by literature studies and then analyzed scientifically, mathematically and statistically for prediction. This research paper exclusively studies the earthquake and its impact on people susceptible to the location. This can somehow specify the magnitude, intensity which can be shared with public for precaution.

An earthquake can be predicted by regression analysis based on its depth, longitude, latitude, location, magnitude etc. Whereas earthquakes can be placed in order of destruction of lives, buildings, injuries, tree uprooted etc. by classification analysis. A set of independent or explanatory variables (P's) with a dependent or response variable(Q) forms a functional relationship and represented as P= Q(X). It incorporates the value/s which is known as predictor variable and values to be predicted as response variable. This paper has indicated the relationship of earthquake data from big data analytics extraction and the occurrences of earthquake. This paper is helpful to recognize the active zones and the frequency of earthquake occurrences based on established relationship on magnitude, depth and location. Three different types of Regression Analysis are shown in (Figure 1).

Footnote

[1] Sumita Mukherjee – Research Scholar and Main Author responsible for Concept and Analysis.

[2] Dr. Prinima Gupta – Co Author and guide for formulation and structuring of the concept.

3 Prof. Felix Musau – Co Author and guide for the application of Regression Analysis.

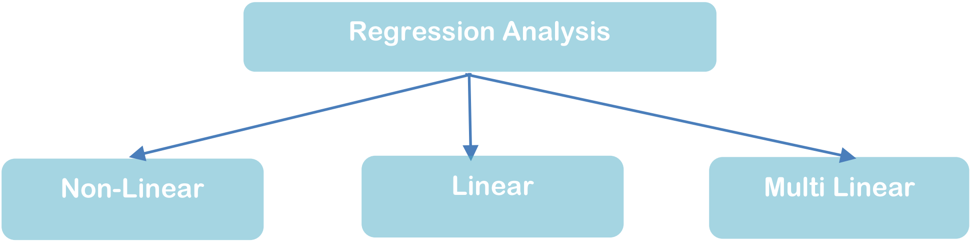


Figure 1*- Regression Analysis*

2. Literature Review

Can there be surety of occurrences of earthquake? How should people overthrow the struggles and troubles encountered. In the study of prediction of earthquake difficulties can be encountered for weather forecasts. The major components for numerical earthquake prediction are of 5 kinds and their current status are concluded along with seismic stations studying seismic waves [1] [2][8]. The major damage and loss were caused in Nepal by earthquake in 2015. For better perception of damage in Nepal by earthquake this paper assesses the seismicity of various sectors based on regions. It inspects the obtained and acquired ground motion data, aftershock data, seism tectonic setting. The possessed damage data in plenty and bounty (geo-tagged photos and observation comments) are assembled using the Kmz file and the Google Earth. These data are made available to public [6][7][11]. During the last decades the occurrences of earthquake and its magnitude is been a demanding aspect where many researchers are working. The approach of researchers is useful to study further and continue and combining robust Mathematical, Geophysical Statistical, and Machine learning techniques to analyze materialized big data having a big dataset [9]. The first priority in the seismic studies is the occurrences of ground motion along with the magnitude. For studying earthquake speeding up of ground, proposed regression models are used which is recorded by seismometers installed at a station in Chiang Mai, Thailand. The recording from 2006 to 2012 of the different 73 earthquakes categorized according to the zones of different regions based on magnitudes and sources of occurrences of earthquake which is grasped by this model [18][20]. For analyzing active fault data, earthquake data multiple regressions are designed, analyzed and then the regression methods are evaluated with Akaike’s Information Criterion (IEEE Trans Autom Control, 19(6):716–723). The AIC method has considered many parameters and best fitted for a regression formula. For calculating the estimation of magnitude by regression analysis this useful formula is applied for occurrences of earthquake [12].

As Mw=1.13log +0.16logR+4.62Mw=1.13log fLs +0.16log R+4.62------------------------------- [1].

Earthquake prediction model selected more than 60 research articles published studied thoroughly and analyzed for better understanding.

Seismic data zone wise based on regions was covered and proposed by the model have endeavored long term predictions regarding location, depth, magnitude, timing, and intensity of future occurrences of earthquakes. This article discusses and involves many different variants of Fuzzy, rule-based, and machine learning based expert systems for earthquake prediction, regression calculation mathematically. Though, rule-based variants include fuzzy, machine learning expert system but applies in different manner [20][21]. In this study a new model which is mathematical in nature was proposed for the tangible contraction strength prediction at different ages and was proposed and developed an equation taking non-linear regression concepts. From the knowledge of the mix itself, i.e. mix proportion the variables, elements are used in the prediction models. This model provides good estimation of compressive strength according to the analysis, including the data used in this study and yields good correlations. For the prediction the correlation coefficients were 0.995 and 0.994 of 7days and 28days compressive strength respectively Moreover, in compressive strength of different concretes predictions, in spite of variations in the results the proposed models proved to be an outstanding tool [12][17].

Data mining plays an important role for prediction of various kinds of disaster. Earthquake is no different. Regression analysis can take both mathematical and statistical form and action that builds a relationship considering two or more flexible parameters in terms of the original units of data. The degree of change in one variable calculates and reflects the change associated with other variable or variables is reflected accurately by Linear regression [3][19]. This model is tested for finding linear relationship occurrence by chance or not! Most of the time there is a puzzlement between two data mining handlers Classification and Regression concepts as both concepts are referred for solving problem of same type like prediction and forecasting. The difference between regression and classification is a continuous value or a numeric value is predicted through regression and classification choses distinct class for data assignment. This regression analysis can be achieved by using Excel or SPSS. Among the anticipating, recognition and characterizing relationship the multiple factors can be represented mathematically and statistically [18][10]. A very popular technique in engineering, social sciences, mathematical is nonlinear regression analysis. The most widely practiced approach to estimation of parameters are parameters estimation Least-squares with Gauss–Newton method. Free hand curve is also used. The reparameterization is inherent to nonlinearity but can be corrected for proper statistical correction [10]. By this study the influence of the complexity of multiple linear regression models on accuracy and software size estimation is investigated also analyzes the significance of variables and UCP (use case points). To analyze the impact of model complexity, stepwise multiple linear regression models and residual analysis were used to study the correlation analysis for the impact of each variable [2][4] [18]. The techniques used for mining of data can also be used for prediction of various normal day to day disasters. This research study the various data mining technique specially the impact and application of regression analysis to predict the occurrences of earthquake. A scrutiny of application of mining of data in the forecast of day to day normal crisis of geological nature is dispensed by this paper [3]. The number of research studies and articles on the subject published between 2013 to 2020 were 18 papers and studied thoroughly for better understanding and application. The data mining techniques mainly used for earthquake predictions are regression models, the Bayesian belief network, and decision trees, time series models, logistic models, neural networks, all of which lead to the problems intrinsic in the prediction of occurrences of earthquakes which are basic and primary explanations to prediction of earthquakes [21].

3. Proposed Methodology

The estimation of motion of ground is the main priority and movement of the seismic waves in the earth's crust ultimately resulting the occurrences of earthquake. Tremors, earthquakes are unpredictable causing death, instability among people's mind and their living. If this study can intimate people one week or month before about the possibility of seismic disturbances and dependency on different parameters like depth, longitude, latitude, magnitude etc.an alternative arrangements can be made. The data gathered for Nepal for five years can be divided into different zones recognizing earthquake prone areas and locations. This paper is aimed at proposing and analyzing regression model like nonlinear, linear and multilinear both mathematically and statistically so that the prediction is accurate. Thus, the result of linear and multilinear are also analyzes significantly to find an accurate correlation of magnitude and depth to the temperature of earth's crust through movement of seismic waves.

3.1 regression analysis: non-linear regression equation

Many different forms are taken in nonlinear equation. To check an equation for its nonlinear properties, the easiest way is to determine above if the criteria is met or not for a linear equation. Nonlinear regression covers the most flexible curve-fitting functionality. Minitab's catalogue provides various examples for nonlinear function. In the nonlinear functions, P represents the predictor or dependent and Q represents the response or independent variable and the other parameters are represented by Thetas. Unlike linear regression, these functions can have more than one parameter predictor variable. The various possible shapes by non linear are drawn (Figure 2.).

Nonlinear regression model of the relationship between Depth and Magnitude considering few years of earthquake in Nepal is shown by an example here. Based on data size the representation by nonlinear equation is too long to fit on graph (Graph 1 and 2).

Table 1*- Earthquake data of Nepal based on Depth and Magnitude*

S.no.	Magnitude (P)	Depth(Q)
1	8.8	93
2	8.7	95
3	8.6	56
4	9.2	64
5	9.7	85
6	6.5	51
7	4.2	23
8	9.1	61
9	9.6	75
10	9.2	59
11	8.5	47

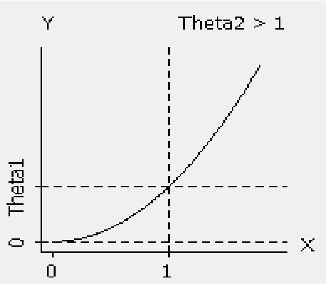
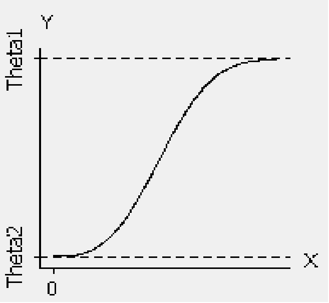
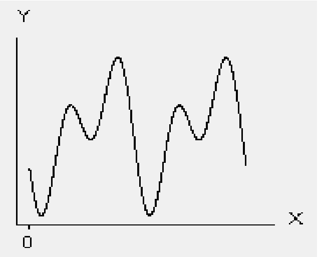
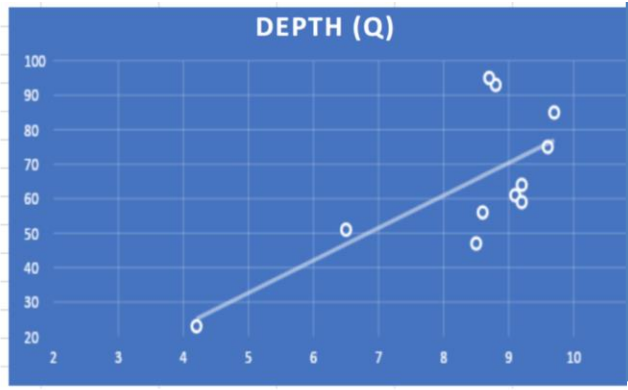
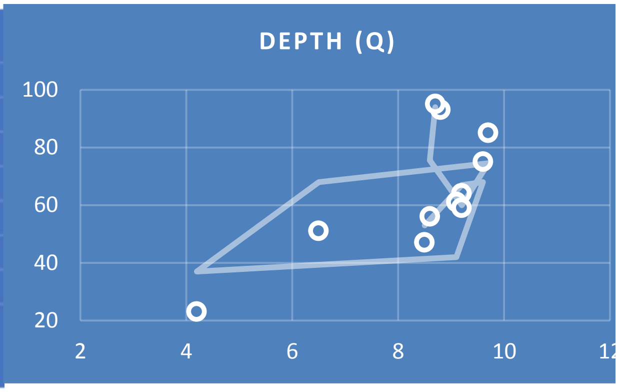
Nonlinear function	*Viable Form (one)*
Power (convex): Theta 1 * X^Theta2	
Increasing Depth: Theta1 + (Theta2 - Theta1) * exp (-Theta3 * X^Theta4)	
Increasing Magnitude: Theta1 * cos (X + Theta4) + (Theta2 * cos (2*X + Theta4) + Theta3	

Figure 2*- Viable Shapes of Non-Linear Equation based on Table-1*



Graph 1- Non -linear graph Depth vs Magnitude



Graph 2- Non-linear graph Depth Vs Magnitude connecting the dots

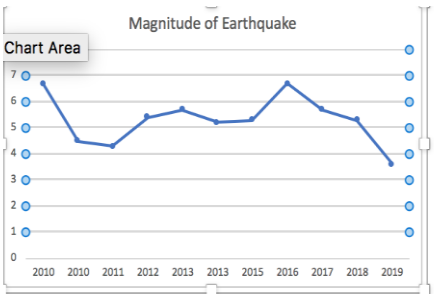
Using the formula here depth can be calculated:
Depth = (1288.14 + 1491.08 * magnitude Ln + 583.238 * magnitude Ln^2 + 75.4167 * magnitude Ln^3) / (1 + 0.966295 * magnitude Ln + 0.397973 * magnitude Ln^2 + 0.0497273 * magnitude Ln^3) ……………. (2)

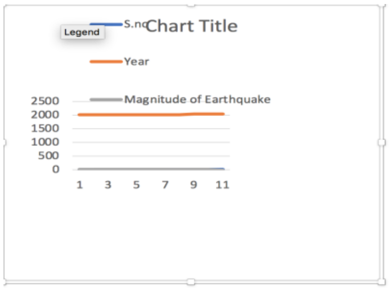
The functional form of the models that each analysis accepts are used actually for naming Linear and nonlinear regression. Non-linear regression is the last attempt specifically in the convergence of the algorithms to calculate the coefficients provided it is not calculated properly by regular regression. When the relationship is not linear between dependent and independent variable, the extension of the linear least square regression of functions is calculated by nonlinear regression for much bigger and general class calculations. When there is no analytical solution for estimating the parameters, a good result is not expected. (see Graph 1,2).

3.1.a Free Hand Curve: A freehand smooth curve is drawn through the plotted points and along with the horizontal axis based on the observation. The curve is drawn in such a manner that the points are plotted on the curve properly with a smooth curve and proper direction. The curve is drawn in such a manner that the concentration is made more on the curve. Once the free hand line or curve is drawn, the estimated values on the Y axis graph can be read for each time period. For segregating the trend this is the fastest and simplest method. This is most suitable to analysis original data producing scattered diagram representing well defined trends. Here X axis represents the year of earthquake in Nepal and y represents the magnitude of the earthquake. We use free hand for nonlinear. It helps us to draw a smooth freehand curve through plotted points based on observed data including the horizontal and vertical axis. Most of the points compress across the curve and removes the irregular movements of data and generates an accepted trend. Only problem to use with large number of earthquake data is no two readings of the curve are same as separate individual draws distinct lines and curves which are individualistic and thus brings variations in interception and slope (Table 2), (Graph 3,4,5).

***Table 2-** Used for Drawing Free hand graph having Year and Magnitude as parameter*

S.no	Year	Magnitude of Earthquake
1	2010	6.7
2	2010	4.5
3	2011	4.3
4	2012	5.4
5	2013	5.7
6	2013	5.2
7	2015	5.3
8	2016	6.7
9	2017	5.7
10	2018	5.3
11	2019	3.6





Graph 3- Free Hand Graph by Excel Data. Graph 4- Free Hand Chart by Excel Data. Graph 5-Free Hand Drawing

3.2 Regression Analysis: Linear Regression Equation The relationship between an independent variable and dependent variable is approached and acquired by Linear Regression. It can be expressed as

$$Q = a + b*P + \delta \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

Where:

Q is Response variable (Dependent)

P is Explanatory variable (Independent)

a is Intercept

b is Slope

δ is Residual value (error)
A unique presentation is followed by linear regression model. A regression model is considered as linear in Statistics when all the components follow the following rules. i) The constant. ii) A variable is multiplied by an independent variable=W. The equation is then formulated by only adding the terms together. These rules limit the form to just one type:

Dependent variable= constant + variable * W+ … + variable * W-----------------------(4)

$Q=\mu_0 + \mu_1 * P_1 + \mu_2 * P_2 + \ldots\ldots + \mu_n * P_n$---(5)

We can get an independent variable by an exponent to fit a curve in case of function being linear with the associated variables. The model can get a U-shaped curve if the independent variables are squared.

$Q= \mu_0 + \mu_1 * P_1 + \mu_2 * P_2$ ^2---(6)

It can be expressed Algebracially for representing mathematically.

3.2a. The Least Square Method: A deciding statistical method known as least square method is used to find a e a best-fit line or a regression line for a given pattern utilizing specific parameters by an equation. This method explains the approximation based on number of equations and provides solution for each equation on different observed data by reducing the sum of the squares of deviations or errors. The value between observed data or the best fit result is analyzed in this model. It is the most suitable and appropriate method for curve fitting. The disadvantage of this method is it might result measurement errors as most of the time curve fitting gives zero errors to independent variable but sometimes it is not true. This method thus forces to have a hypothesis testing for error free result. We can calculate a predictive model that can let us estimate magnitude far more accurately through the magic of least sums regression, than by sight alone and with a few simple equations. Regression analysis is the extremely powerful analytical tool, used within technology and science. To construct complicated regression models for a large requirement there are a number of popular statistical programs used for an appropriate solution. A calculator is a simpler model, such as this requires nothing more than some data. At this point this method is intended for continuous data which is not worth.

Least Squares Regression Equations

The assumption of a regression model is to examine the impact of one or more variables (in this case injuries occurred during earthquake) are independent in nature (in this case magnitude of earthquake) on a dependent

variable of earthquake. Linear regression analyses such as these are based on a simple equation:
Q= a+ b*P---------------------------(7) (3.2a)

Here Q is depth occurred during earthquake, a is the intercept, b is the coefficient or slope and P is the magnitude of the earthquake

The regression equation of least square is used to forecast or predict the feedback of Q for an individual value of P which is descriptive in nature. This regression line is produced as small as possible by adding the squares of the distances of vertical direction of the observed data points. This method is used mathematically as it is adjusted and applied directly or indirectly to many situations with a wider range and to other modelling methods, to get a satisfactory result. This method treats continuous quantity as residuals and also predicts the changes of input to the degree of output changes in a function through derivatives. Sometimes it does not generate dependable test statistics as in place of absolute value of the offsets square value of offsets are used thus effecting the distance points of the line than closer ones. (Table 3 is used for calculation).

Table 3*- For Least Square Method Calculation Mathematically*

S. No	Magnitude. (P)	Depth (Q)	Magnitude- Average Magnitude (P-P □)	Depth- Average Depth (Q-Q □)	(P-P □) ^2	(P-P □) * Q-Q □)
1	8.8	93	0.4	28.5	0.16	0.060
2	8.7	95	0.3	30.5	0.09	-0.231
3	8.6	56	0.2	-8.5	0.04	0.014
4	9.2	64	0.8	-0.5	0.64	-0.677
5	9.7	85	1.3	20.5	1.69	-1.031
6	6.5	51	-1.9	-13.5	3.61	0.000
7	4.2	23	-4.2	-41.5	17.64	13.441
8	9.1	61	0.7	-3.5	0.49	-0.322
9	9.6	75	1.2	10.5	0.24	1.550
10	9.2	59	0.8	-5.5	0.64	0.550
11	8.5	47	0.1	-17.5	0.01	0.787
Total	92.1	709	-0.3	17.5	25.2	
Mean	8.4 (P □)	64.5 (Q □)				

$$b = \sum(P - \bar{P}) * (Q - \bar{Q}) / \sum(P - \bar{P})^2$$

b= (-0.3*17.5)/25.2 b=--5.25/25.2 b=-0.20---------------(8)

Q= a+b*P

64.5=a+(-0.20) *8.4

64.5=a+-1.68

a=-64.5-1.68   a= 66.18--(9)

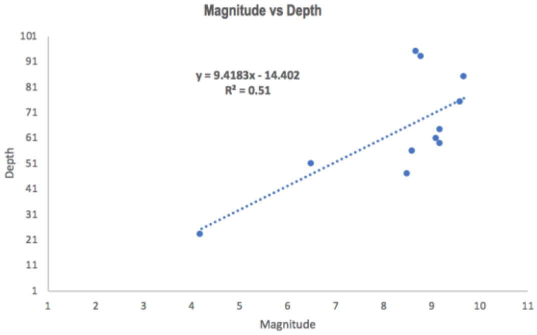
Q=a+b*P

Q=66.18+(-0.20) *P Now we can put any value of P to calculate Q for example if P

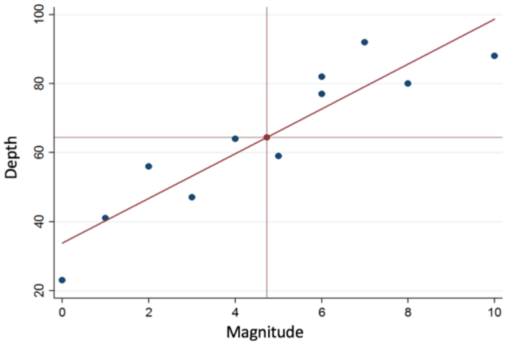
is 8.8 Q will be

Q=66.18+(-0.28) *8.8 Q= 66.18-2.46
Q=63.72

To draw a least square regression line by hand for best fit the estimated depth for a series of magnitude cane be connected through ruler. This line will cross the means of depth and magnitude. The graph in Excel is also plotted by using the same data (Graph 6 & 7).



Graph 6- Plotted in Excel



Graph 7- Drawn Free   Hand

3.2.b. Deviation from the Arithmetic mean method: When the values of P and Q are large the least square method application becomes clumsy. The deviation from the arithmetic mean method becomes the most appropriate and simple to handle the large values. This method derives two formulas for Regression Equations.

***Table 4-** Calculating Arithmetic Mean Method Mathematically*

S. No	Magnitude. (P)	Depth (Q)	i=Magnitude-Average	j= Depth-	i*i	j*j	i*j

			Magnitude (P-P □)	Average Depth (Q-Q □)			
1	8.8	93	0.4	28.5	0.18	814.8	12.2
2	8.7	95	0.3	30.5	0.11	933.0	10.0
3	8.6	56	0.2	-8.5	0.05	71.5	-1.9
4	9.2	64	0.8	-0.5	0.68	0.2	-0.4
5	9.7	85	1.3	20.5	1.76	422.1	27.3
6	6.5	51	-1.9	-13.5	3.51	181.0	25.2
7	4.2	23	-4.2	-41.5	17.41	1718.5	173.0
8	9.1	61	0.7	-3.5	0.53	11.9	-2.5
9	9.6	75	1.2	10.5	1.51	111.2	12.9
10	9.2	59	0.8	-5.5	0.68	29.8	-4.5
11	8.5	47	0.1	-17.5	0.02	304.7	-2.2
Total	92.1	709	-0.33	0.00	26.4	4598.7	249
Mean	8.4	64.5					

The standard error of the regression is taken to assess the precision of the predictions thus it plays a vital role like R squared. When we use regression model 95% approximately should lie between plus or minus of data observed, from the regression line standard error of the regression *2 is expected (Table 4 used for calculation)

a. Equation of regression of P on Q --→ $(P- \bar{P}\ ) = R_{ij}\,(Q - \overline{Q)}$

b. Equation of regression of Q on P -→ $(Q\ -\overline{Q)}\ = R_{ji}\,(P- \bar{P})$

R_{ij} and R_{ji} are regression coefficients

$R_{ij.} = \frac{\sum i*j}{\sum j^2}$ and $R_{ji} = \frac{\sum i*j}{\sum i^2}$

Regression equation of P on Q=

(P- P □)=R_{ij} (Q- Q □)

$R_{ij} = \sum (I *j)/ \sum I *j$

P-8.4=249/4598.7(Q-64.5)

P-8.4=0.1(Q-64.5)
P=0.1Q +1.95--(10)

Regression equation of Q on P=
(Q-Q □) = Rji (P-P □)

$R_{ji=} \sum(i* j)/ \sum i* i$

Q-64.5=249/26.4(P-8.4)

Q=9.4P-14.46---(11)

RV=1*8.8+2=10.8 The residual value Rv is 8.8-10.8= -2.0-------------(12)

3.2.c. Deviation from Assumed Mean Method: It simplifies calculating accurate values by hand and is used to quickly estimate statistical calculations. Assumed mean is the assumption of the mean is a simple means to calculate true mean and standard deviation. It does not give accurate results on large observations. Here 4.2 is assumed as the mean of P and 23 as the mean of Q (Table 5 used for calculation).

Table- 5 *For Calculating Assumed Mean Method Mathematically*

S.no.	Magnitude. (P)	Depth (Q)	Di	Dj	di*di	dj*dj	di*dj
1.	8.8	93	4.6	70	21.1	4900	322
2.	8.7	95	4.5	72	20.2	5184	324
3.	8.6	56	4.4	33	19.3	1089	145.2
4.	9.2	64	5.0	41	25.0	1681	205
5.	9.7	85	5.5	62	30.2	3844	341
6.	6.5	51	2.3	28	5.2	784	64.4
7.	4.2	23	0.0	0	0	0	0
8.	9.1	61	4.9	38	24.0	1444	186.2
9.	9.6	75	5.4	52	29.1	2704	280.8
10.	9.2	59	5.0	36	25.0	1296	180
11.	8.5	47	4.3	24	18.4	2209	103.2
Total	92.1	709	45.5		217.5	23502	2151.8
Mean	8.4	64.5					

P □= ∑P/N=92.1/11=8.4

Q □= ∑Q/N =799/11=64.5

The Regression Coefficient of P on Q Rij = (N* $\sum d_{i*}\, d_j$ - $\sum d_i$- $\sum d_j$)/N-$\sum d_i^2$-(d_j) ^2

=11*2151.8-45.9*526=-473.6   11*23502-526*526=- 18154 R_{ij}= -473.6/-18154=0.026

P- $\overline{P}$= R_{ij} $(Q - \overline{Q})$

P-8.4= 0.026* (Q- 64.5)

P=0.026Q-64.5+8.4

P=0.026Q - 56.1---(13)

The Regression Coefficient of Q on P R$_{ji}$ = $\frac{N * \sum d_i * d_j - \sum d_i * \sum d_j}{N * \sum d_i^2 - (\sum d_i)^2}$

=11*2151.8-45.9*526=-473.6 11*218-45.9*45.9=291.19=-1.62

$Q - \overline{Q}$ = Rij (P- $\overline{P}$)

Q-64.5=-1.62*(P-8.4)

Q-64.5=-1.62P+13.6

Q=-1.62P+64.5+13.6

Q=-1.62P + 78.1

Q=1.62P – 78.1---(14)

RV=Sd/√N=-8.27/3.3=-2.5--(15)

Table 6- Comparative analysis with Residual value

S.No.	The regression of P on Q	The regression of Q on P	Residual value (RV)
SD Using Arithmetic Mean	P=0.1Q +1.95	Q=9.4P-14.46	-2.0
SD using Assumed Mean	P= 0.026Q - 56.1	Q=1.62P – 78.1	-2.5

This shows the calculation on both methods are quite close to each other showing accuracy.

3.3. Regression Analysis: Multiple linear regression

There is hardly a significant difference between multiple linear regression and simple linear regression except that the usage of independent variables is many in nature. It can be expressed as

B = p + qA$_1$ + rA$_2$ + dA$_3$ + ϵ --(16)

Where:

B – Dependent variable

A_1, A_2, A_3 – Independent (explanatory) variables

p – Intercept

q, d – Slopes

ϵ – Residual (error)

Though the value for equation mathematically can be calculated but historical data of earthquake of Nepal is used so that the result can be calculated statistically by using ANOVA Table in Excel. Regression Analysis is performed for the analysis of earthquake in Nepal. The analysis examines the relationship between location and depth and magnitude of an earthquake, where in magnitude of an earthquake is dependent on its depth and location.

The approach was to divide the data of each country into 4 zones – North, South, West and East and used the excel Data tab > Data analysis feature to calculate regression for each zone of the location and we performed regression for the overall data as well. While calculating regression, Depth and location were taken as the independent variables and Magnitude as the dependent variable. It could find that the depth and location of earthquakes and the temperature of the earth's crust. This analysis is capable of identifying the outliers, or anomalies. For example, while reviewing the data related to earthquake prediction the researcher can find the longitude, latitude, time, magnitude correlated to depth and location. The frequency of Magnitude and Depth, was also examined, taking median as base for our analysis (Table- 7,9,10,11).

3.3.a Statistical Analysis overall zones: For Statistical analysis a sample of data in Excel is taken and applied Regression Analysis through ANOVA Table.

***Table 7-** Earthquake Data of Nepal of an appropriate segment is used from Big Data*

S. No	Date	Time (UTC)	Time (IST)	Longitude	Latitude	Depth	Magnitude	Location	Province	Zone
1	31/01/2015	13:59:43	19:29:43	28.37°N	84.07°E	10	5.0	Pokran, Nepal	4	East
2	22/01/2015	3:42:00	9:12:00	29.42°N	81.06°E	45	7.0	Nepal	6	North
3	05/01/2015	19:41:00	1:11:00	29.17°N	81.63°E	10	4.6	Nepal	6	North
4	30/04/2015	0:37:01	6:07:01	28.01°N	84.65°E	10	4.2	Kathmandu, Nepal, Gorakhpur, India	3	East
5	29/04/2015	11:27:05	16:57:05	27.88°N	85.51°E	14	4.8	Northwest of Nagarkot	3	East
6	29/04/2015	17:16:00	22:46:00	27.8505°N	85.6739°E	10	4.3	Pokhran, Kathmandu	3	East
7	27/04/2015	23:20:00	4:50:00	27.8836°N	85.1996°E	10	4.3	Gokarneshwor, BiratNagar	3	East
8	27/04/2015	21:27:00	2:57:00	27.7555°N	85.6771°E	10	4.3	ENS of Nagarkot	3	East
9	27/04/2015	21:27:00	2:57:00	27.7555°N	85.6771°E	10	4.3	ENE of Nagarkot	3	East
10	27/04/2015	21:14:00	2:44:00	27.68°N	85.25°E	10	3.9	Nepal, Kanpur	3	East

11	27/04/2015	18:59:00	0:29:00	28.15°N	84.75°E	19	4.8	Garh, Bhawanipur, Udaynarayanpur	4	East
12	27/04/2015	15:51:00	21:21:00	27.73°N	85.14°E	2	2.5	Bhaktapur, Imadol, Kathmandu	3	East
13	27/04/2015	14:57:00	20:27:00	27.824°N	85.9208°E	10	4.1	Jaynagar, Kathmandu	3	East
14	27/04/2015	12:35:00	18:05:00	26.85°N	88.09°E	21	5.0	Kathmandu, Noida, India	3	East
15	26/04/2015	14:57:00	20:27:00	28.16°N	84.66°E	2	2.7	Kathmandu, Pokhran	3	East
16	25/04/2015	10:53:00	16:23:00	27.7719°N	85.8701°E	10	4.4	Birpara, Nepal	1	South
17	25/04/2015	10:40:00	16:10:00	27.9°N	85.93°E	10	4.8	Kathmandu, Nepal	3	East
18	25/04/2015	10:23:00	15:53:00	27.8732°N	85.762°E	10	4.2	West of Kodari, Nepal, Munger	3	East
19	25/04/2015	9:30:00	15:00:00	27.8732°N	85.762°E	10	4.8	Nepal, Kolkatta, Bhandavgargh, India	3	East
20	25/04/2015	8:55:00	14:25:00	27.66°N	85.62°E	10	5.0	Nepal, Kharaghpur, Pondicherry	3	East

***Table 8-** Frequency Analysis based on Earthquake Data of Nepal*

Frequency Analysis					
Magnitude	Overall	North	West	South	East
Less than 4.2	155	3	0	6	146
More than 4.2	270	14	0	12	244

Depth	Overall	North	West	South	East
Less than 10	28	3	0	1	24
More than 10	397	14	0	17	366

3.3.b Statistical Analysis North Zone

***Table 9-** Earthquake Data extracted of North Zone*

S. no	Date	Time(UTC)	Time (IST)	Longitude	Latitude	Depth	Magnitude	Location	Province	Zone
1	22/01/2015	3:42:00	9:12:00	29.42°N	81.06°E	45	7.0	Nepal	6	North
2	05/01/2015	19:41:00	1:11:00	29.17°N	81.63°E	10	4.6	Nepal	6	North
3	21/04/2015	14:02:00	19:32:00	28.89°N	82.4°E	27	5.2	Nepal	6	North
4	22/05/2015	19:07:23	0:37:23	29.8°N	81.74°E	2	2.4	Gangabu, Kiritpur, Kathmanndu, Nepal	6	North
5	18/11/2015	08:25:21	13:55:21	29.8°N	80.5°E	15	4.3	NEPAL-INDIA BORDER REGION	6	North
6	18/12/2015	22:16:10	03:46:10	29.34°N	81.68°E	10	4.2	Patan, Nepal, Delhi, India	6	North
7	18/12/2015	22:16:03	03:46:03	28.71°N	81.48°E	9	5.1	Nepal	6	North
8	29/06/2	09:10:	14:40	29.55°	81.28°	15	4.8	Nepal	6	North

	016	18	:18	N	E					
9	06/11/2017	03:09:17	08:39:17	29.5824°N	81.2208°E	10	4.5	NE of Dipayal, Nepal	6	North
10	14/04/2019	22:42:12	04:12:12	29.6463°N	81.4054°E	29.1	5.0	62km NE of Dipayal,Nepal	7	North

3.3.c Statistical Analysis East Zone

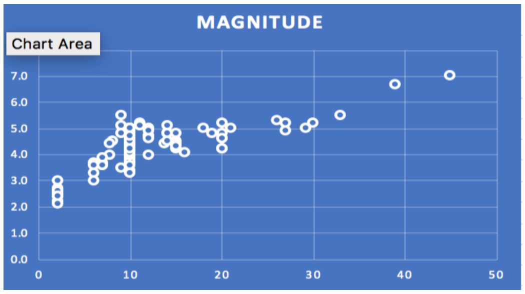
***Table 10**- Earthquake Data Extracted of East Zone*

S.no	Date	Time(UTC)	Time(IST)	Longitude	Latitude	Depth	Magnitude	Location	Province	Zone
1	31/01/2015	13:59:43	19:29:43	28.37°N	84.07°E	10	5.0	Pokran, Nepal	4	East
2	30/04/2015	0:37:01	6:07:01	28.01°N	84.65°E	10	4.2	Kathmandu, Nepal, Gorakhpur, India	3	East
3	29/04/2015	11:27:05	16:57:05	27.88°N	85.51°E	14	4.8	Northwest of Nagarkot	3	East
4	29/04/2015	17:16:00	22:46:00	27.8505°N	85.6739°E	10	4.3	Pokhran, Kathmandu	3	East
5	27/04/2015	23:20:00	4:50:00	27.8836°N	85.1996°E	10	4.3	Gokarneshwor, BiratNagar	3	East
6	27/04/2015	21:27:00	2:57:00	27.7555°N	85.6771°E	10	4.3	ENS of Nagarkot	3	East
7	27/04/2015	21:27:00	2:57:00	27.7555°N	85.6771°E	10	4.3	ENE of Nagarkot	3	East
8	27/04/2015	21:14:00	2:44:00	27.68°N	85.25°E	10	3.9	Nepal, Kanpur	3	East
9	27/04/2015	18:59:00	0:29:00	28.15°N	84.75°E	19	4.8	Garh Bhawanipur, Udaynarayanpur	4	East
10	27/04/2015	15:51:00	21:21:00	27.73°N	85.14°E	2	2.5	Bhaktapur, Imadol, Kathmandu	3	East

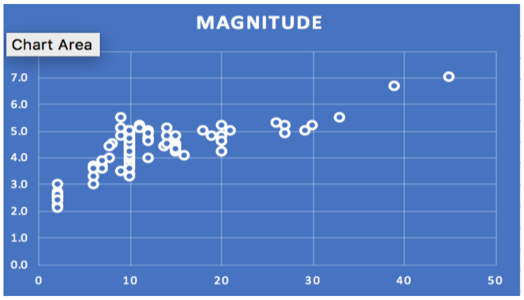
3.3.d Statistical Analysis South Zone

***Table 11**-Earthquake Data extracted of South Zone*

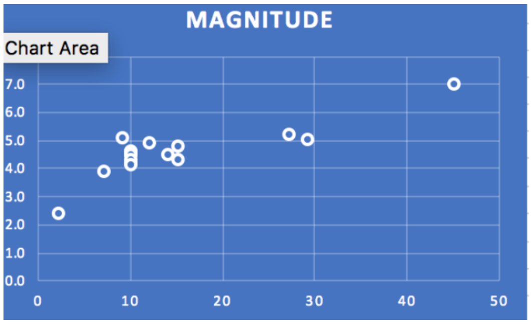
S.no	Date	Time (UTC)	Time (IST)	Longitude	Latitude	Depth	Magnitude	Location	Province	Zone
1	25/04/2015	10:53:00	16:23:00	27.7719°N	85.8701°E	10	4.4	Birpara, Nepal	1	South
2	26/05/2015	14:15:18	19:45:18	27.54°N	85.43°E	10	4.3	Biratnagar, Bharatput, Surat, India	2	South
3	26/05/2015	8:26:48	13:56:48	27.63°N	86.3°E	10	4.3	Baratput, Patan, Nepal, Raxaul	2	South
4	24/07/2025	7:10:10	12:40:10	27.8°N	86.29°E	10	4.4	Biratnagar, Bharatput, Silguri, India	2	South
5	24/07/2025	1:21:14	6:51:14	27.67°N	86.18°E	10	4.2	Baratpur, Nepal Motihari, India	2	South
6	21/05/2015	8:21:07	13:51:07	27.85°N	86.32°E	10	4.2	Nepal	1	South
7	13/05/2015	18:31:30	00:01:30	27.92°N	86.31°E	10	4.3	Nepal	1	South
8	12/05/2015	22:53:05	04:23:05	27.82°N	86.47°E	10	4.0	Madhyapur Thimi, Nepal	1	South
9	12/05/2015	17:28:11	22:58:11	27.66°N	86.31°E	10	4.3	Nepal,Kolkatta,Lucknow ,India	1	South
10	12/05/2015	07:48:41	13:18:41	27.59°N	86.54°E	10	5.0	Nepal, UP,Kolkatta	1	South



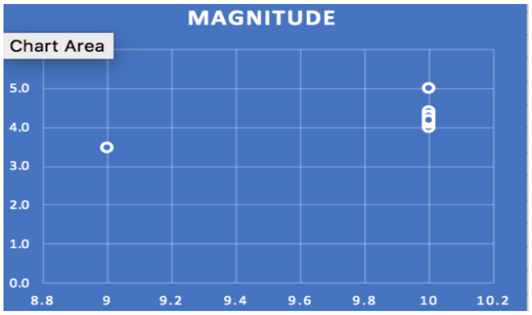
Graph 8- For Overall Earthquakes in Nepal.



Graph 9- Earthquakes in Nepal of East zone



Graph 10- Earthquakes in Nepal North Zone

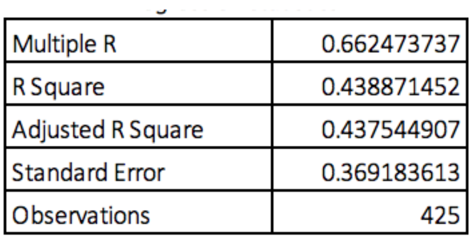


Graph 11- Earthquakes in Nepal South Zone

4. Result and Discussion

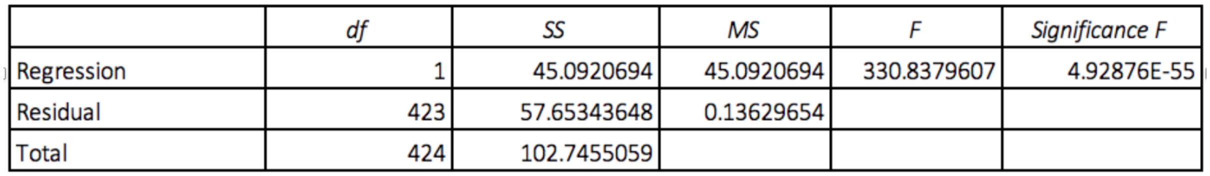
The data of Nepal of five years are filtered from big data. The significant data are tabulated according to necessary parameters like Date, Time, Longitude, Latitude, Depth, Magnitude, Location etc. The frequency table is created based on two parameters Depth and Magnitude. The location is also considered to categorize according to four zones- North, East, South and West. West Zone is not considered as there is not adequate data for earthquake occurrences. Research has been carried out, by using the data, for earthquake prediction through interchange of earthquake data based on seismology observations and these data were collected for five years from Nepal. A regression model of data mining is developed to study the probability through different attributes. Thus, it can be observed that by using the following algorithmic model for earthquake prediction, proper methods can be implemented for deploying warnings and preparing for earthquakes (Table 8).

***Table 12**- Analysis of Regression Statistics of overall earthquake occurrences of Nepal*



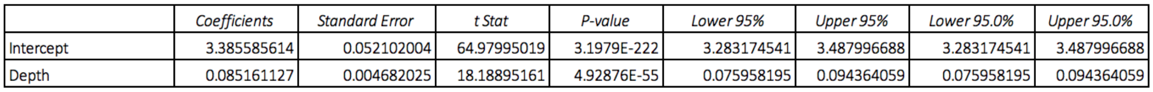
Multiple R	0.662473737
R Square	0.438871452
Adjusted R Square	0.437544907
Standard Error	0.369183613
Observations	425

Table 13- *table for overall representation of earthquake of Nepal*



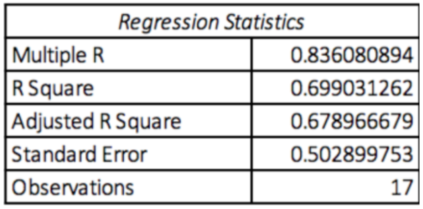
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	45.0920694	45.0920694	330.8379607	4.92876E-55
Residual	423	57.65343648	0.13629654		
Total	424	102.7455059			

Table 14- *Parameter estimates based on overall representation of earthquakes of Nepal*



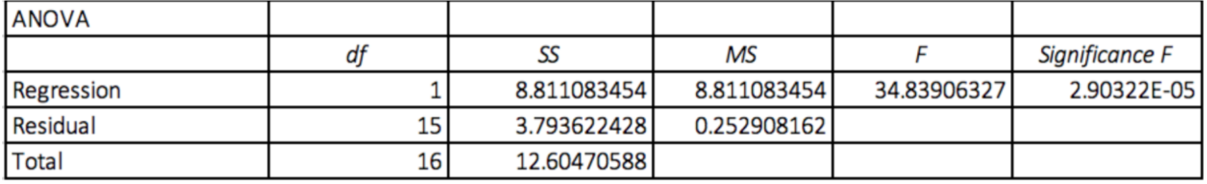
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.385585614	0.052102004	64.97995019	3.1979E-222	3.283174541	3.487996688	3.283174541	3.487996688
Depth	0.085161127	0.004682025	18.18895161	4.92876E-55	0.075958195	0.094364059	0.075958195	0.094364059

Table 15- *Analysis of Regression Statistics for earthquakes of North zone of Nepal*



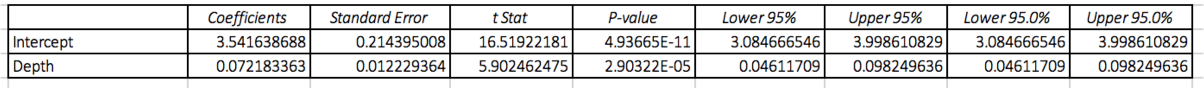
Regression Statistics	
Multiple R	0.836080894
R Square	0.699031262
Adjusted R Square	0.678966679
Standard Error	0.502899753
Observations	17

Table16- *ANOVA table for earthquakes of North zone of Nepal*



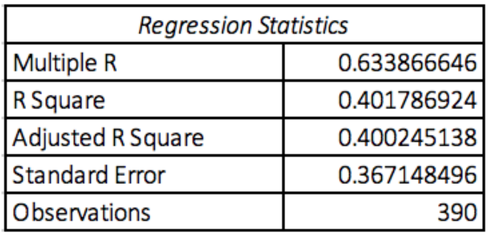
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	8.811083454	8.811083454	34.83906327	2.90322E-05
Residual	15	3.793622428	0.252908162		
Total	16	12.60470588			

Table 17- *Parameter estimates based on earthquakes of North zone of Nepal*



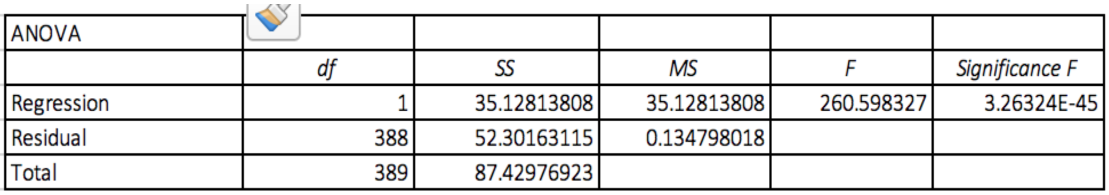
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.541638688	0.214395008	16.51922181	4.93665E-11	3.084666546	3.998610829	3.084666546	3.998610829
Depth	0.072183363	0.012229364	5.902462475	2.90322E-05	0.04611709	0.098249636	0.04611709	0.098249636

Table 18- *Analysis of Regression Statistics of earthquakes of East zone of Nepal*



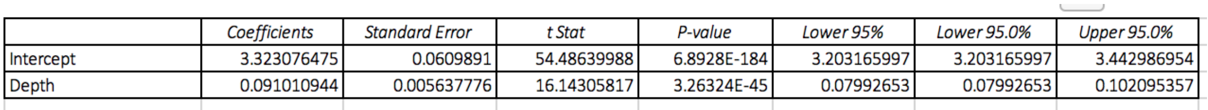
Regression Statistics	
Multiple R	0.633866646
R Square	0.401786924
Adjusted R Square	0.400245138
Standard Error	0.367148496
Observations	390

Table 19- *ANOVA table for earthquakes of East zone of Nepal*



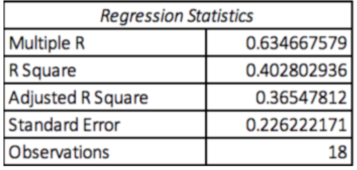
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	35.12813808	35.12813808	260.598327	3.26324E-45
Residual	388	52.30163115	0.134798018		
Total	389	87.42976923			

Table 20- *Parameter estimates based on earthquakes of East zone of Nepal*



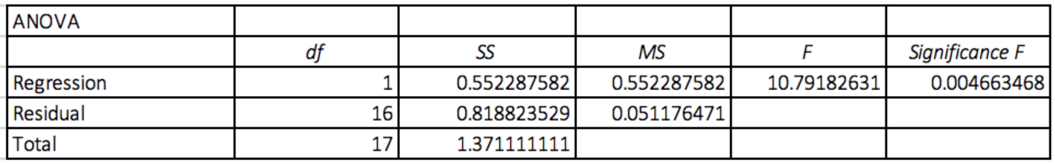
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.323076475	0.0609891	54.48639988	6.8928E-184	3.203165997	3.203165997	3.442986954
Depth	0.091010944	0.005637776	16.14305817	3.26324E-45	0.07992653	0.07992653	0.102095357

Table 21- *Analysis Regression Statistics of Earthquakes of South zone of Nepal*



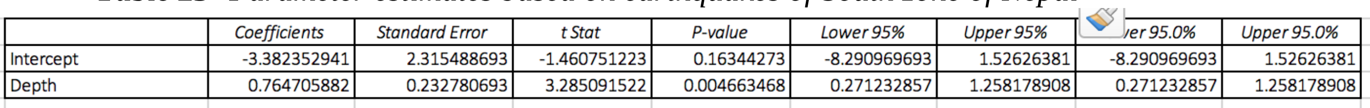
Regression Statistics	
Multiple R	0.634667579
R Square	0.402802936
Adjusted R Square	0.36547812
Standard Error	0.226222171
Observations	18

Table 22- *ANOVA table for earthquakes of South zone of Nepal*



ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.552287582	0.552287582	10.79182631	0.004663468
Residual	16	0.818823529	0.051176471		
Total	17	1.371111111			

Table 23- *Parameter estimates based on earthquakes of South zone of Nepal*



	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ver 95.0%*	*Upper 95.0%*
Intercept	-3.382352941	2.315488693	-1.460751223	0.16344273	-8.290969693	1.52626381	-8.290969693	1.52626381
Depth	0.764705882	0.232780693	3.285091522	0.004663468	0.271232857	1.258178908	0.271232857	1.258178908

Table 24-*Comparative R Squared in different Zones*

Location Based Data	Multiple R	R Square	Adjusted R Square	Standard Error	Observations
Earthquakes of Nepal of all zone for 5 years	0.662	0.438	0.437	0.369	425
Earthquakes of North Zone	0.836	0.699	0.678	0.502	17
Earthquakes of East Zone	0.633	0.401	0.400	0.367	390
Earthquakes of South Zone	0.634	0.402	0.365	0.22	18

Residual plots or values created from statistical analysis through ANOVA Table can depict a biased model far more effectively than the numeric or mathematic output showing by displaying unappropriated patterns. The scatter of the data points is decided by R squared across the fitted regression line. In multiple regression the range of R squared values between 40 to 60 percent between fitted and observed data is considered good for prediction.

At shallower depth of layer of crust of earth there is an occurrence of larger earthquakes whereas the tiny and minor earthquakes do and can take place at all depths. The earthquakes having less than magnitude 7, i.e. small scaled magnitudes are known to occur in both the crust and the slabs, which are subducting in nature. The earthquakes having magnitudes up to the largest observed magnitude of 9.5 (1960 Chile earthquake), i.e. large scaled magnitude, typically take place within the crust of the earth. These greater and larger earthquakes are only predominant at cool and unfriendly temperatures and are associated with frictional sliding on faults at layers of crust of the earth (Graph 8, 9,10,11).

The simple answer is that the major and big earthquakes occur at shallower depths in the earth's crust, but minor and small earthquakes can and do occur at all depths down to about 700 km (400 mi).

The layer of the earth, which is considered topmost typically 7 to 30 km (4 to 18 mi) thick is the platform where the earthquakes occur in the layer of the crust of the earth. The crust has many fault systems and is the most fragile, coldest part of the earth on which earthquakes occur. Frictional sliding on the faults on earth's crust by the buildup tectonic stress causes these earthquakes to occur.

5. Conclusion

Once the earthquake data using parameter like longitude, latitude, depth, magnitude, location etc. of Nepal were collected from the meteorological and seismological department studied thoroughly then the regression method is applied. The regression method which is applied both mathematically and statistically can increase the accuracy of predicting the magnitude and depth related to a location based on a probable zone. The overall R square was between 40% -50% for Nepal suggests that the larger earthquakes occur at flat and hollow depths in the layers of crust of earth and that to in a particular zone. But tiny and minor earthquakes can and do take place at all depths and varied zones of different regions. Those earthquakes that have magnitude less than about 7, are known as earthquake having small magnitude and are viable to occur both in the crust and the subducting slabs (Table 16). Those earthquakes having magnitude up to the largest observed magnitude of 9.5 (1960 Chile earthquake), are known as major earthquake typically occur within the layer of crust of the earth. These huge, broad and major earthquakes are associated with frictional sliding on faults, which can only occur at cooler temperature.

This study also indicates that the occurrences of earthquakes has a tendency to occur at hollow flat depth of earth's crust, whereas smaller earthquakes can appear and occur at all depths of the earth's crust. The earthquakes which are considered smaller having a lesser magnitude than 7 are predicted to occur in both crust and subducting slabs of the earth. The earthquakes having larger magnitude mostly occur at the location where temperature is consistently cold and only in the last layer of earth's crust due to fictional sliding of faults as well as at the regions having colder temperature. This research also justifies that zone wise occurrences of earthquake are dependent of temperature, magnitude and depth of earth's crust.

Declaration Sumita Mukherjee is been working hard with metrology and seismology department of India and Nepal to collect the authentic data for five years. Studied thoroughly the data mining analysis, its concept, application and usage for prediction. Firstly, I would like to thank my guide Dr. Prinima Gupta, and co -guide Prof. Felix Musau for their guidance and support. I can't thank you both enough for the constant support and guidance, discussion and help to solidify my ideas to write this paper. I would like to extend my thanks to my dear colleagues Ms. Carolyne, Ms. Maryanne Gichuhi, and Mr. Faithful Wachira for co- operating and motivating me. I also grateful to two Children of mine Ms. Sunaina Backaya and Mr. Supratik Mukherjee for your service.

References

MFM Zain, Suhad Mabd, Mathematical regression model for the prediction of concrete strength Proceedings of the 10th WSEAS international conference on Mathematical methods, computational techniques and intelligent systems, Pages 396–402, 2008.

Astrid Schneider, Gerhard Hommel, Prof.Drrernat Linear Regression Analysis, Evaluation of Scientific Publications, Deutsches Arzteblatt International Journal, DoI:10.3238/arztebi.2010.0776, 2010.

Shrey K. Shahi & Jack W. Baker Department of Civil and Environmental Engineering Stanford University, Stanford, CA, USA Regression models for predicting the probability of near-fault earthquake ground motion pulses, and their period, 2010.

Seiya Uyeda, Masashi Kamogawa, Toshiyasu Nagao Short term earthquake prediction: Current state of seismo-electromagnetics in Tectonophysics, DOI: 10.1016/j.tecto.2008.07.019470(3-4):205-213, 2010.

Francisco Martinez-Alvarez, Alicia Tronciso, Antonio Morales, Computational Intelligence Techniques for Predicting Earthquakes Conference Paper DOI: 10.1007/978-3-642-21222-2_3, 2011.

Grantej Vinod Otari, R.V. Kulkarni, A Review of Application of Data Mining in Earthquake Prediction, Geography, Corpus ID: 18062442, 2012.

Chorin, A and M orzf eld, M. Conditions for successful data assimilation. J. Geophys. Res. - At mos. 118, 11522-[5]11533. (Cross ref), (Web of Science ®). [Google Scholar) ,2013.

Yaolin Shi, Bei Zhang, Siqi Zhang, On numerical, Earthquake prediction, Earthquake Science, volume 27, pages319–335, 2014.

lkram, A, &amp; Qamar, U. A rule -based expert system for earthquake prediction. Journal of Intelligent Information. Systems, 43(2), 205-230. doi: 10.1007/s10844-014-0316-5 [Cross ref), [Web of Science ®). (Google Scholar), 2014.

E. Florido, F. Martinez, Alvarez, A. Morales- Esteban, J., Reyes, and J. L Aznarte. Detecting precursory patterns to 406 enhance earthquake pre diction in Chile. Computers and Geosciences, 76:112- 120, 2015.

Katsuichiro Goda, Ram Mohan Pokhrel, Gabriele Chiaro, The 2015 Gorkha Nepal Earthquake: Insights from Frontiers in Built Environment DOI:10.3389/fbuil.2015.00008, 2015.

Mehdidoust, J. Z., &amp; Shahbahrami, A Study of Expert Systems in Predicting Earthquake. (Google Scholar), 20 Journal of Earth and Space Physics, volume 42, Page(s) 281 To 292, 2016.

Festim Halili, Avni Restami, Predictive Modeling: Data Mining Regression Technique Applied in a Prototype International Journal of Computer Science and Mobile Computing, Vol.5 Issue.8, pg. 207-215,2016.

Earthquake Prediction in California Using Regression Algorithms and Cloud-based Big Data Infrastructure Computers &Geosciences DOI: 10.1016/j.cageo.2017.10.011, 2017.

M orzfeld, M., Hodyss, D., &amp; Snyder, C. What the collapse of the ensemble Kalman filter tells us about particle filters A: Dynamic Meteorology and Oceanography, 69(1), 1283809. doi: 10.1080/16000870.2017.1283809 [Taylor & Francis Online], [Web of Science ®)] [Google Scholar], 2017.

Khushbu Kumari, Sunil Yadav Linear regression analysis study Department of Anthropology, University of Delhi, New Delhi, India Cross Ref DOI: 10.4103/jpcs.jpcs_8_18, 2018 vector regressor and hybrid neural networks methodology, Published online July 5. doi: 10.137 l / journal. pone .0199004, 2018.

Ghorban i, S.; Barari, M.; Hoseini, M. Presenting a new met ho d to improve the detection of micro -seismic events. Environment Assess. 190,464. [Cross Ref], 2018.

Vasti, M.; Dev, A. Classification and Analysis of Rea l-World Earthquake Data Using Various Mach in e Learning Algorithms. In Lecture Note s in Electrical Engineering; Springer: Singapore, pp. 1- 14, 2019.

Arnaud Mignan, Marco Broccardo Neural Network Applications in Earth quake Predict ion (1994 -2019):

Meta- Analytic In sight on their Limitations, University of Liverpool, October, 2019.

Asim, K.M.; Moustafa, S.S.; Niaz, I.A.; Elawad i, E.A.; Iqbal, T.; M artinez -Alvarez, F. Seismicity analysis and machine learning mode ls for short -term low magnitude seismic activity predictions in Cyprus. Soil Dyn. Earthq. Eng. 130, 105932. [Cross Ref], 2020.

Rabia Tehseen, Muhamma d Shoaib Farooq, Adnan, Earthquake Prediction Using Expert Systems: A Systematic Mapping Study Department of Computer Science, University of Management and Technology, Lahore 54770, 2020