iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** V **Month of publication:** May 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

To Predict Rain Fall in Desert Area of Rajasthan Using Data Mining Techniques

Peeyush Vyas
Asst. Professor, CE/IT Department of Vadodara Institute of Engineering, Vadodara

Abstract: ***Weather forecasting has always been one of the most challenging problems in the desert areas of Rajasthan. Rainfall has become a significant and technical factor in the desert state of India like Rajasthan. Data mining is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and*** **systems. There** ***are various forecasting methods available for the prediction of rainfall and among these methods data mining techniques can be effectively used in predicting rainfall, humidity and wind pressure. Some of the techniques are called as Empirical techniques and others are called as called dynamic techniques. In the empirical method, historical data of the rainfall is compared and processed with the available climatic variables of different parts of the globally available data. Regression, Artificial Neural Network, fuzzy logic etc. are the most widely used techniques under empirical approaches. Whereas, in dynamic approach predictions are created by physical model and is implemented using numerical rainfall forecasting method. Clustering, classification and Multiple Linear Regression are used in this paper for rainfall prediction.***
Keywords: Data mining, Classification, Clustering, Regression, Multiple Linear Regression, Desert Area, Weather forecasting.

I. INTRODUCTION

A. Climatic Conditions of Rajasthan

There is a large and arid region in the western part of India, known as “Thar Desert” of Rajasthan. It is the largest and biggest desert of India which creates a natural boundary between India and its neighbor country Pakistan. The climate of Rajasthan state varies from arid to sub-humid. The average rainfall in Western Rajasthan is below 35 cm and in eastern Rajasthan it is below 75 cm in a year. In summer, temperature goes around 49 degrees and in winter it fall up to zero degree in some of the places. This desert covers about 70% of the land which is bounded by “Aravali mountains” in the east and “Satlej river” in the north-west. Infertile Rajasthan, which spreads in twelve western districts of the State covering about 20 million ha, faces very worst situation due to almost every year drought and low rainfall. The climate of Rajasthan state varies from arid to sub-humid. Rainfall plays a vital role in evaluation and management of drought especially in the desert areas of Rajasthan. Most of the parts of Rajasthan depend on the rainfall for their agriculture needs.

B. Rainfall Forecasting

Rain forecast is basically done by gathering quantitative data about the existing status of the atmosphere on a given place and using scientific understanding of atmospheric processes to plan how the atmosphere will develop on that place. Prediction of rainfall in advance definitely helps in proper agricultural planning. Also, it helps in planning to make necessary arrangements for procurement, transport and distribution of food grains if there is below normal rainfall. The objective of this paper is to seek the factors that can be responsible for the prediction of the rainfall in the desert area, using some of the techniques of Data Mining. Techniques like Classification, Clustering, Multiple Linear Regression, Artificial Neural Network etc. can be effectively used to study the above said problem.

C. Data Mining Techniques

Data mining is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. It involves database and data management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data Mining techniques like Classification, Clustering, Association Rule

mining, Text Mining and Multiple Linear Regression can be effectively used to study all the above said problems. This research does not depict any planned association as we usually find with these data mining techniques but still these techniques provide some association of conditions that can be used as warnings for future disaster.

D. Multiple Linear Regressions

Regression is a data mining technique used to fit an equation to a dataset. Linear regression is the simplest form of regression thatuses the formula of a straight line as –

$$y = mx + b \quad \text{.....(1)}$$

Here, y = Dependent variable
x = independent variable
m, b =regression parameters

It determines the appropriate values for m and b to predict the value of y based upon a given value of x.
It is an statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the relationship between the explanatory and response variables.

Themodel for MLR, given n observations, is:

$Y_i = B_0 + B_1x_{i1} + B_2x_{i2} + ... + B_px_ip + E_i$ where i = 1,2, ..., n

MLR takes a group of random variables and tries to find a mathematical relationship between them. The model creates a relationship in the form of a straight line (linear) that best approximates all the individual data points.
Correlation and regression analysis are related in the sense that both deal with relationships among variables. The correlation coefficient is a measure of linear association between two variables. Values of the correlation coefficient are always between -1 and +1. A correlation coefficient of +1 indicates that two variables are perfectly related in a positive linear sense whereas, a correlation coefficient of -1 indicates that two variables are perfectly related in a negative linear sense. A correlation coefficient of 0 indicates that there is no linear relationship between the two variables. For simple linear regression, the sample correlation coefficient is the square root of the coefficient of determination, with the sign of the correlation coefficient being the same as the sign of b1, the coefficient of x1 in the estimated regression equation.

The quantity r, called the linear correlation coefficient measure the strength and direction of relationship between the two variables. Mathematically r can be defined as -

$$r = \frac{n\sum(xy) - \sum(x)\sum(y)}{\sqrt{(\sum(x2) - \sum(x)^2) - \sqrt{(\sum(y2) - \sum(y)^2)}}}$$

It measures how well the regression line represents data, if the regression line passes through every point on the scattered plot it would be able to explain all of the variation.

R-squared = Explained variation / Total variation

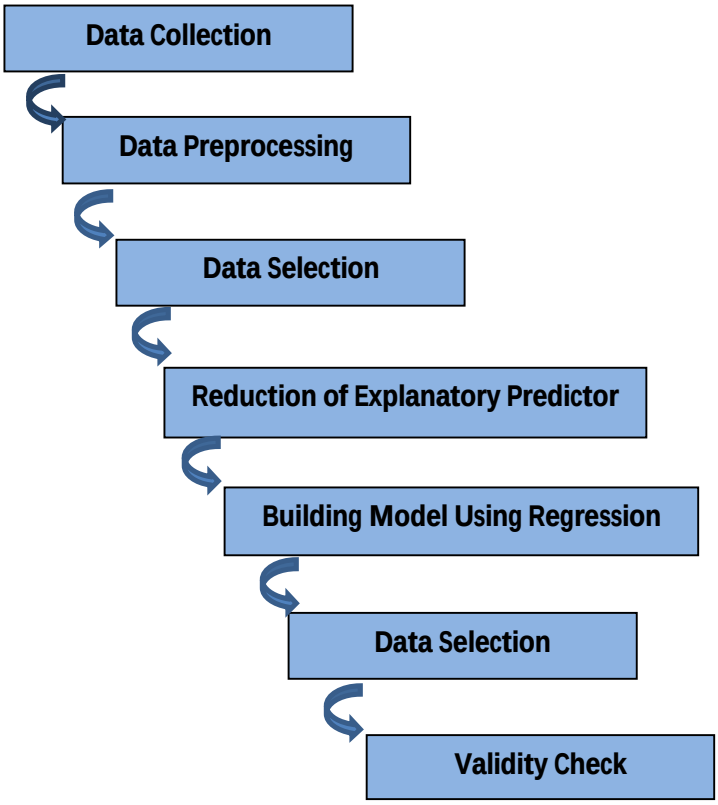
A high $\mathbf{r^2}$ shows that there exists a linear relationship between the two variables. If $\mathbf{r^2=1}$, it indicates the perfect relationship between the two variables. The standard error of the estimate is a measure of the variability of predictions in a regression. Let us consider *yest* as the estimated value of y for a given value of x. This estimated value can be obtained from the regression curve of y on. From this, the measure of the scatter about the regression curve is supplied by the quantity:

$$Sy.x = \sqrt{\frac{\sum(y - yest)2}{n}}$$

The above equation is called the Standard Error of Estimate of y on x.

E.Rainfall Prediction Using Multiple Linear Regression

There are various methods included in forecasting rainfall like Data Collection, Data selection and preprocessing, Reduction of explanatory predictor, building model using regression and finally validity check.
Data Collection is the first most important step for data mining .The Weather dataset is collected from Hydro met Division of Indian metrological department. The department maintains the dataset on monthly basis.
The next step is Data Preprocessing in data mining. The obtained data may have some missing values and may be noisy, so, it is to be cleaned by filling the missing values and deleting some useless data.
The third step is Data selection. Useful data is selected here which are relevant for analysis. Then, we use correlation to determine which data are correlated and which are or not.
The existence of highly inter correlated variables may increase the sampling variation of the regression coefficients and may reduce the model analytical capacity. So, the predictors which have greater inter correlation with others are reduced.
In the next step, linear regression technique is used for Building model with the use of training data. Finally, validity is checked.



F. Training Data set of Jaisalmer and its district

HYDROMET DIVISION, NEW DELHI
INDIA METEOROLOGICAL DEPARTMENT
DISTRICT RAINFALL (mm) FOR LAST FIVE YEARS

District : JAISALMER

Note :

(1) The District Rainfall in millimeters (R/F) shown below are the arithmetic averages of Rainfall of Stations under the District.

(2) % Dep. are the Departures of rainfall from the long period averages of rainfall for the District.

(3) Blank Spaces (if any) show non-availability of Data.

Year/Month	May		June		July		August		Sept.		October	
	R/F	%DEP.	R/F	%DEP.	R/F	%DEP.	R/F	%DEP.	R/F	%DEP.	R/F	%DEP.
2009	0.0	-100	23.6	32	40.4	-27	19.0	-64	5.5	-70	0.0	-10
2010	0.0	-100	85.6	378	88.2	60	124.0	134	62.9	238	1.7	7
2011	3.1	-53	2.7	-86	58.9	-1	121.7	114	104.2	357	0.0	-10
2012	3.2	-52	9.6	-51	17.0	-71	96.5	70	86.4	279	0.1	-9
2013	2.0	-70	29.3	51	43.8	-26	120.4	112	5.7	-75	3.6	1

G. Experimental Results

The various experiments are performed to evaluate the accuracy of rainfall prediction using multiple linear regression. The prediction results are reported in this section. To measure the quality of the MLR equation, the predicted rainfall amount is compared with actual rainfall. For experiments, regional rainfall data taken from Jaisalmer division and western Rajasthan-India and precipitation, cloud cover, average temperature and wind taken fromhttp://www.indiawaterportal.org/articles/district-wise-monthly-rainfall-data-2004-2010-list-raingauge-stations-india-meteorological retrieved on 02/03/2014 [8]. are used as predictors. The data set for 30 years is used for the experiment. The following table shows the details of the predictor's correlation with the rainfall for prediction

Table1: Correlation of predictor with rainfall

Sno.	Predictor	Correlation Coefficient
1	Precipitation	0.8948
2	Cloud Cover	-0.5465
3	Average Temperature	0.9813
4	Wind	-0.4242

In the regression analysis, there are different types of approach like- linear regression, log based, and nonlinear regression for prediction. Here, we have used multiple regression approach on the data set. From this approach we can predict rainfall in any one of the future's year by using climatic factors. Now, for moving towards this approach first we select 4 climate factors with rain dataset of Jaisalmer city, Rajasthan, India .we apply multiple regression approach on that data set and find out predictable equation between rain and climate factors .So, MLR is given below.

Y=-1323.062+0.237*X1+0.75*X2+16.317*X3+13.038*X4

Where Y=Predicted rainfall

X1= Precipitation

X2= Average Temperature

X3= Cloud Cover

X4= Wind

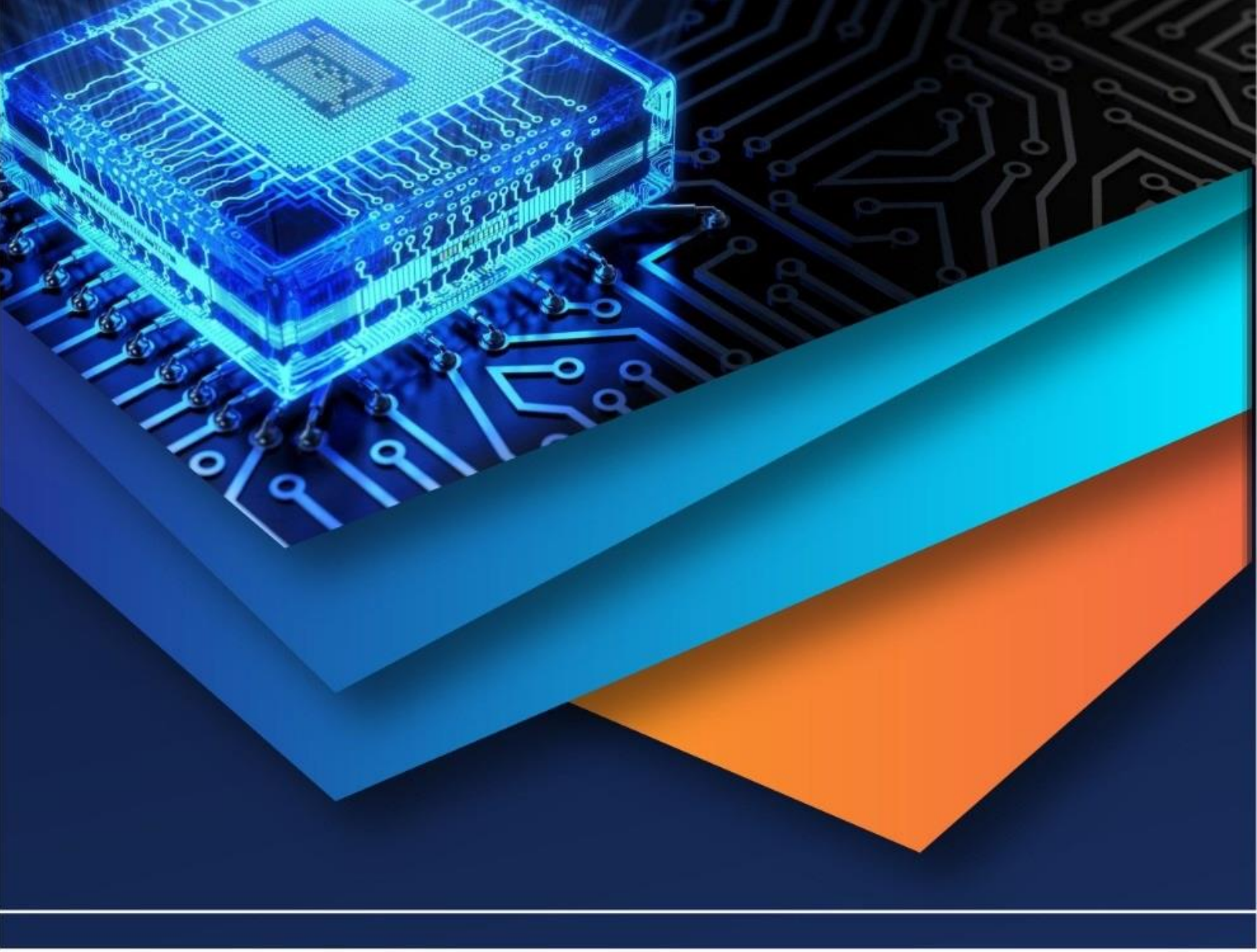
From this equation, we can calculate Rainfall for future years by knowing the precipitation, average temperature, cloud cover and Wind. When the MLR equation is used with test data for testing the accuracy of the MLR equation we obtain the rainfall amount which is close to the actual rainfall data.

II. Conclusion

The topic of monsoon-rainfall data series is highly complex and the multiple linear regressions might play a vital role for the future research. Predictive model can be very useful for the future monsoon. This regression method can be used to forecast rainfall in any area. In this paper, a method for rainfall prediction is proposed after the analysis of Jaisalmer(Western Rajasthan) data set which is derived using correlation analysis and regression analysis. This type of prediction can be very useful for farmers and agriculture purpose. This prediction regarding the rain may not be very much accurate due to the climatic factors. Still, some challenges are there in the prediction of rainfall in the desert areas so, better data mining techniques can be implemented for the rain fall prediction. in which better implementations of data mining technique can be carried out in the field of weather forecasting.

REFERENCES

[1]Neha Khandelwal et.al " Climatic Assessment Of Rajasthan's Region For Drought With Concern Of Data Mining Techniques" in International Journal Of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 5, September- October 2012, pp.1695-1697.
[2] Paras ,et.al," A Simple Weather Forecasting Model Using Mathematical Regression" in Indian Research Journal of Extension Education Special Issue (Volume I), January, 2012.
[3] Ozlem Terzi ,"Monthly Rainfall Estimation Using Data-Mining Process" in Hindawi Publishing Corporation Applied Computational Intelligence and Soft Computing Volume 2012, Article ID 698071, 6 pagesdoi:10.1155/2012/698071.
[4] Z.ismail,et.al," Forecasting Gold Prices Using Multiple Linear Regression Method" in American Journal of Applied Sciences 6 (8): 1509-1514, 2009 ISSN 1546-9239.
[5] Wint Thida Zaw ,et.al." Empirical Statistical Modeling of RainfallPrediction over Myanmar in World Academy of Science, Engineering and TechnologyVol:22 2008-10-270.
[6]S. Nkrintra, et al., "Seasonal Forecasting of Thailand Summer Monsoon Rainfall", in International Journal of Climatology, Vol. 25, Issue 5, American Meteorological Society, 2005, pp. 649-664.
[7]http://blog.minitab.com/blog/adventures-in-statistics/regression- analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit retreieved on 24/04/2014.
[8]http://www.indiawaterportal.org/articles/district-wise-monthly-rainfall-data-2004-2010-list-raingauge-stations-india-meteorological retrieved on 02/03/2014.
[9] http://www.biochemia-medica.com/content/standard-error-meaning- and-interpretation retrieved on 25/04/2014.
[10] H.Hasani,et al," A New Approach to Polynomial Regression And Its Application to Physical growth of Human Height.
[11] http://cs.gmu.edu/cne/modules/dau/stat/regression/multregsn/mreg_2 _frm.html retrieved on 25/04/2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY