American Journal of Power Electronics and Power Systems
1(1): 12-15, August 2019

Research Paper

Precipitation Prediction Using Modified Linear Regression.

[1]Dr. Anjali Saxena, [2]Vandana Barya, [3]Dr. Mamta Singh

[1,2,3.]Asst Professor, Dept. of Mathematical Science & Computer Applications, Bundelkhand University, Jhansi, Utter Pradesh, India.

To cite this article: Dr. Anjali Saxena, Vandana Barya, Dr. Mamta Singh. Precipitation prediction using modified linear regression. American Journal of Chemstry and Chemical Engineering, 1(1): 12-15,May-June 2019.

Email: anjali_saxena3@yahoo.com

Received: 14th March 2019.|Revised: 28th May 2019.|Accepted: 14th June 2019.

Abstract: Investigation frequently includes considering previous recorded information to inquire about potential patterns. Climate condition is the condition of environment at a given time as far as climate factors like precipitation, cloud conditions, temperature, and so forth. The current models utilize information mining systems to foresee the precipitation. The principle burden of these frameworks is that it doesn't give a gauge of the anticipated precipitation. The framework ascertains normal of qualities and comprehend the condition of air, which doesn't yield gauge results. This paper speaks to a numerical strategy called Linear Regression to foresee the precipitation in different areas in southern conditions of India. The Linear Regression strategy is altered keeping in mind the end goal to acquire the most ideal mistake rate by repeating and adding some level of blunder to the information esteems. This technique gives a gauge of precipitation utilizing distinctive barometrical parameters like normal temperature and overcast cover to foresee the precipitation. The straight relapse is connected on the arrangement of information and the coefficients are utilized to foresee the precipitation dependent on the comparing estimations of the parameters. The fundamental preferred standpoint of this model is that this model gauges the precipitation dependent on the past relationships between the diverse climatic parameters. In this way, a gauge estimation of what the precipitation could be at a given era and place can be found effortlessly. Catchphrases: precipitation forecast, linear regression.

Keywords: Rainfall, Linear regression, multiple linear regression, dataset, Test data.

1. Introduction

The utilization of science and innovation that predicts the condition of environment at some random specific day and age is known as Weather determining. There is a wide range of strategies to climate gauge. Climate estimate sees are critical in light of the fact that they can be utilized to counteract demolition of life and condition. The climate determining techniques utilized in the old time more often than not inferred design acknowledgment i.e., they as a rule depend on watching examples of occasions. For instance, it is discovered that the next day has brought reasonable climate; if the first day dusk is especially red. In any case, the majority of the forecasts end up being dependable. Here in this framework we utilized parameters like normal temperature, overcast cover to foresee the precipitation. The informational index of 100 years is taken for this task and the actualized utilizing Numerical strategies. Climate anticipating framework utilizes environmental parameters like dampness, wind and temperature and figure climate dependent on going before record, in this manner, this determining is more solid. There are numerous application that this framework be utilized, for example, Air Traffic, Agriculture, Marine, Forestry, Navy, and Military and so on. The preparation of the information is finished utilizing an adjusted form of direct relapse strategy. The mistake rate between the genuine and anticipated is utilized to enhance the preparation set and prepare the information with the new data sources Thus, with the preferences this technique, we can estimate the results till there is no further change in the blunder rate. This technique can be utilized to conjecture the precipitation and keep the obliteration caused by it to the life or property.

2. Background

2.1. Linear Regression

It is a method used for defining the relation between a dependent variable (Y) and one or more independent variables or explanatory variables, denoted by (X). For multiple explanatory variable, the process is defined as Multiple Linear Regression (MLR). The general equation for a linear regression is given as

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \cdots \ldots \ldots + \beta_\varphi x_{i\varphi} + \varepsilon_i = x_i^\tau \beta + \varepsilon_i,$$
$$i = 1 \ldots \ldots \ldots . n$$

Where y_i denotes the dependent variable (rainfall) and xi where i=1, 2... n, denotes the explanatory or independent variables and β is called the intercept.

The general linear regression equation used in this system is given as

$$\text{Rainfall} = (\text{AvgTemp} * \beta 1) + (\text{Cloud Cover} * \beta 2) + \beta 3$$

Where $\beta 1$, $\beta 2$, $\beta 3$ represents the different coefficients for different districts.

2.2. Related work

Daniela Şchiopu **[1]** and his team in his publication used SPSS 13.0 tool and forecasted temperature from data collected from the Hong Kong Observatory website. They used factor analysis technique in the SPSS tool to reduce the complexity in calculations the temperature using correlation and regression.

Samuel and Raajalakshmi **[2]** used multiple linear regression to predict the monsoon rainfall by using outgoing long wave radiations, global temperatures and sunspots out of Tamil Nadu. They collected data from 110 years from Indian Meteorological department, Chennai.

Hirani and Nitin **[3]** proposed different methods to estimate rainfall. The methods include Autoregressive Integrated Moving Average (ARIMA), Multiple Linear Regression (MLR), Genetic Algorithm, Support Vector Machine (SVM), Back-Propagation Neural Network(BPNN), Adaptive Splines Threshold Autoregressive (ASTAR) modelling and others.

Paras and Sanjay **[4]** built up an estimating model utilizing scientific relapse. The climate information is gathered for a time of 3 years and this model can foresee max and min temperatures for a time of 15 to 45 weeks into what's to come.

Goutami **[5]** utilized Multiple Linear Regression to assess normal summer – storm precipitation on the information from 1871 to 1999. She investigated the month to month precipitation of Indian summer rainstorm months.

Kannan, Ramachandran and Prabhakaran **[6]** executed multiple straight relapse and Karl Pearson coefficient. They made a short – term gauge over a specific state. They utilized fluffy sets, neural systems to examine the information.

Retius and Delson **[7]** built up a weighted various relapse display. They utilized mix of time arrangement investigation and relapse to offer a ground-breaking framework for foreseeing yearly precipitation.

Timothy and Shukla **[8]** proposed the F-test and Screening system. a cross-approval method is utilized first to screens models out that are for the most part liable to poorly perform on autonomous datasets, at that point the blunder of each model is contrasted with those every single other model with decide edge of centrality in mistake fluctuation.

Guhathakurta **[11]** utilized unique models on nonlinear conditions that air framework oversees. They actualized neural systems with three layer that takes a shot at one info, one yield and one concealed layer. The system preparing is done till mean square mistake of 0.0005 to 0.001.

Nikhil **[12]** proposed relationship and relapse both direct and numerous straight relapse. He evaluated the precipitation by investigating the barometrical variables like precipitation, vapor weight, normal temperature and overcast cover.

2.3. Data collection

The data necessary for the system to predict rainfall are previous year rainfall data, average temperature and cloud cover over the particular area. The data is collected for a period of 70 years i.e., from 1901 to 1970 for each month. This data is then used to process in the system to predict the rainfall. The atmospheric data is collected from the website http://www.indiawaterportal.org/met_data/ published by Arghyam initiative. In this the data for various parameters and states is collected to train the system. The data is collected from the user regarding the state, district, and the year and month which he wants to predict the rainfall. This data is used to formulate the equation for predicting the rainfall by calculating the average temperature and cloud cover at that particular area.

3. Methodology

We use a modified version of Linear Regression to perform the prediction of rainfall in our system. The process of this method is explained in this following steps

1. The input data sets are examined. The input data of training set is obtained from 1901 to 2002 for each month to perform the proposed system and check the method.

2. The training and test data are formed from the input data sets. The training set contains the average temperature, rainfall and cloud cover from 1901 to 1970 from the input data sets. The proposed method is applied on this training sets. The test data contains the data from 1971 to 2002 on which the testing of model is done.

3. The linear regression is applied on the training data sets and the rainfall is forecasted using the rainfall in training data as dependent variable and average temperature and cloud cover as independent variables.

4. The error percentage is now calculated by subtracting the predicted value from the actual value and multiplying it with 100 to get percentage.

5. The error percentage, we add a certain percentage of error percentage to the input training set and repeat the steps 3 and 4 till there is no further increase in the error. The iterative steps that need to be followed are given by

a. Include 20 percent of mistake rate to the info preparing precipitation esteem and now these qualities are utilized in the direct relapse strategy to prepare i.e., stages 3 and 4 is executed.

b. In the event that the outcomes got are not fulfilled, we can rehash the stage 3, 4, 5 till the point where by including the mistake rate there will be no change in the anticipated precipitation.

c. We can attempt the above advance by expanding the blunder rate esteem that is added to the preparation set, i.e., say we are not happy with the mistake rate after the main emphasis, so we can build the mistake rate that is included the following cycle to 40 % and increment so on..

d. In the above advances the coefficients of the free factors in the direct relapse continue evolving.

6. the most recent refreshed coefficients to figure the test information and this delivers the most exact gauge esteems.

Table-1: Input training data sheet of Krishna District, August month

Year	Rainfall	Average Temperature	Cloud cover
1901	120.387	27.88	70.716
……..	……	…….	…….
1966	126.49	29.074	69.242
1967	90.936	27.997	71.668
1968	28.831	29.353	69.242
1969	90.497	28.273	69.799
1970	198.853	27.291	71.421

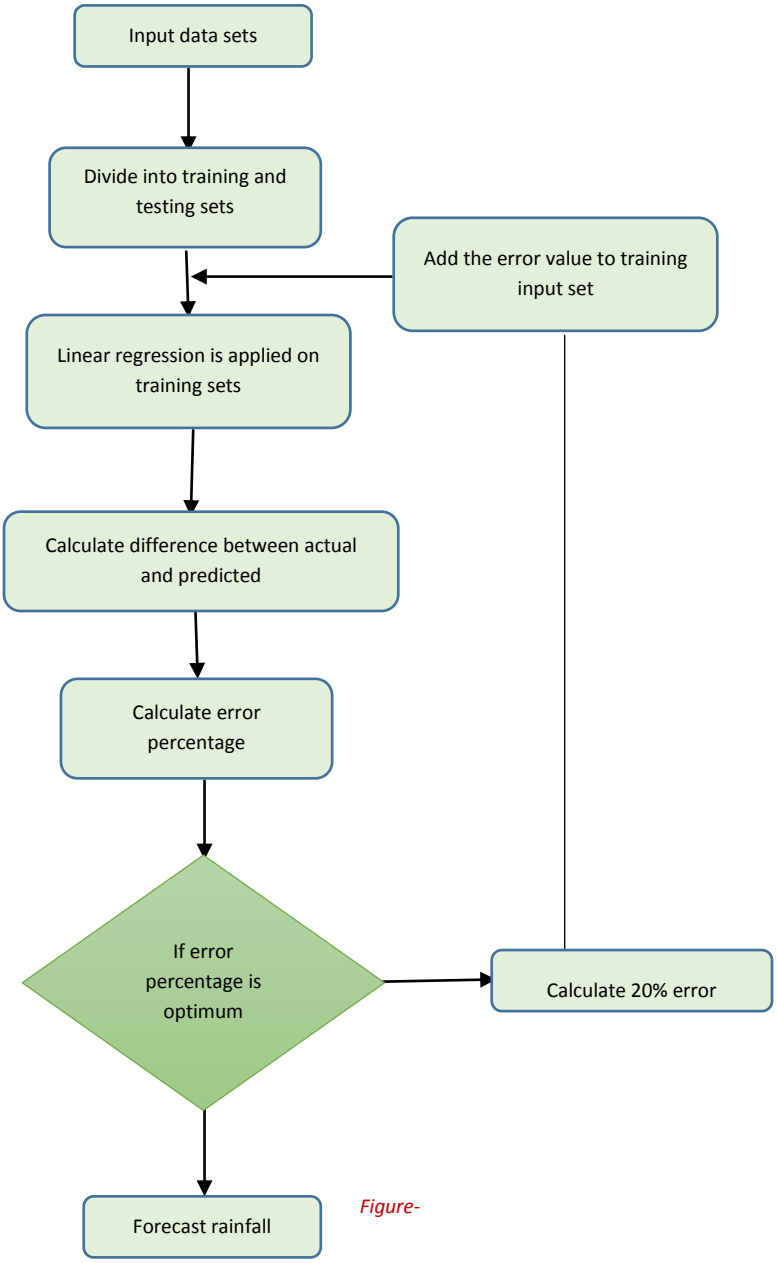

Figure-1: System Architecture

The input data is used to compare the state and district along with month and year, so that the right coefficients are used to perform the forecast, i.e., the values are used in calculating the rainfall. The system automatically find the relative coefficients based on the user input of state, district, year and month. And performs the necessary calculations and provide the forecast values in the output page.

4. Experimental Results

The following is the experimental usage of the system proposed. The experiment is carried out on the Andhra Pradesh State – Krishna District. The data is collected for every district for a period of 102 years from 1901 to 2002 for all the months i.e., from January to December. The training is done using the data from these years. There are four columns in the input year, rainfall, average temperature, cloud cover. The rainfall is taken as dependent variable as it the variable that we are trying to predict, the average temperature and cloud cover are the independent variables that are used to predict the rainfall over the future. The example sample input data is shown in the Table-1, which is used as training data set and perform linear regression on it.

Table-2: Details of new values of regression parameters changing.

	Average Temperature coefficient	Cloud cover coefficient	Constant value
First iteration	-53.378	0.647	1584.936
Final iteration	-52.785	0.598	1604.693
Times performed	5	5	5

So as to acquire the normal temperature and overcast cover estimation of future era the direct relapse is connected to average temperature and year to get the condition to figure the normal temperature, correspondingly for the overcast cover. The accompanying outcomes are gotten after the fifth cycle.

Table -3: Forecast Result

Year	Actual Rain Fall	First predicted Value	Final predicted rainfall	Error percent
1973	146.344	133.811857	136.869892	6.47
1974	122.028	118.990924	122.453906	-0.34
1975	96.152	125.195051	128.491728	-33.63
1976	176.213	168.273058	171.229017	2.82
1977	150.176	159.603798	162.478115	-8.19

The average error percentage of the above test is around 7%, the negative sign indicates that the error is decreased overall.

5. Conclusion

Precipitation is the significant reason for a large number of the cataclysmic events like glimmer surges, dry spells, waves. So to keep these regular disasters, we ought to have the capacity to foresee the reason for the source. The proposed framework can be utilized to evaluate the precipitation over the required period with the goal that the particular experts can avoid potential risk to keep the death toll and property. The proposed framework utilizes adjusted direct relapse way to deal with foresee the precipitation that has less blunder rate than contrasted with most information mining methods like grouping, back proliferation which gives the summed up qualities as opposed to evaluate esteems. This information is utilized to play out the important counts to anticipate the precipitation from normal temperature and overcast front of that specific area. We may enhance this framework additionally utilizing different relapse which can take various months on end as information and simply shaping a solitary condition which drives closer to an exact precipitation anticipated. The proposed approach may likewise be utilized in different applications like, in schools to foresee the normal characteristics of their understudies, in games to anticipate the scores or winning groups dependent on their past execution, in ventures to assess their benefits, and so forth…

References

1. Daniela Şchiopu, Elia Georgiana Petre and Catalina Negoina. Weather Forecast using SPSS Statistical Methods. Whether predictions, 111(39): 100-116, 2010.
2. R.Samuel Selvaraj and Raajalakshmi. Statistical Method of Predicting the Northeast Rainfall of Tamil Nadu. Universal Journal of Environmental Research and Technology, 1(4):557-559.
3. Dhawal Hirani and Dr.Nitin Mishra. A Survey on Rainfall Prediction Techniques. International Journal of Computer Application, 6(2):235-239, March- April 2016.
4. Paras and Sanjay Mathur. A Simple Weather Forecasting Model Using Mathematical Regression. A Master's thesis submitted to Department of Electronics & Communication Engineering, College of Technology, G.B. Pant University of Agriculture & Technology, Pantnagar, (India) 263 145.
5. Goutami Bandyopadhyay. The Prediction of Indian Monsoon Rainfall. A Regression Approach" 1/19 Dover Place Kolkata-700 019 West Bengal India.
6. M.Kannan, S.Prabhakaran and P.Ramchandran. Rainfall Forecasting Using DataMining Techniques. International Journal of Engineering and Technology , 2(6):397-401, 2010.
7. Delson Chikobvu and Retius Chifurira. A Weighted Multiple Regression Model to Predict Rainfall Patterns -Principal Component Analysis approach. Mediterranean Journal of Social Sciences, MCSER Publishing, Rome-Italy.
8. Timothy DelSole and J.Shukla. Linear Prediction of Indian Monsoon Rainfall. Journal of statistics and mathematics, 6(4):276-283, 2011.
9. India Metrological department (1973). Northeast monsoon; FMU Rep.no .iv:18.4.
10. S.Balachandran, R.Asokan and S.Sridharan. Global surface temperature in relation to Northeast monsoon rainfall over Tamil Nadu. J. Earth. syst. Sc., 115:349-362, 2006.
11. P.Guhathakurta. Long-range monsoon rainfall prediction of 2005 for the districts and subdivision Kerala with artificial neural network. Current Science, 90:773-779, 2005.
12. Nikhil Sethi: Exploiting Data Mining Technique for Rainfall Prediction. International Journal of Computer Science and Information Technologies, 5(3):3982- 3984, 2014.
13. P.Guhathakurta. Long-range monsoon rainfall prediction of 2005 for the districts and subdivision Kerala with artificial neural network. Current Science, 90:773-779, 2005.
14. S.Hastenrath. Prediction of Indian Monsoon Rainfall - Further Exploration. Journal of Climate, 1:298-304, 1998.
15. D.C.Montgomery and A.J.Lynwood. Forecasting and Time Series Analysis(text book). First edition, New York, McGraw-Hil, 1996.