VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

# RAINFALL PREDICTION USING MODIFIED LINEAR REGRESSION

S. Prabakaran, P. Naveen Kumar and P. Sai Mani Tarun
Department of Computer Science and Engineering, SRM University, Kattankulathur, Chennai, India
E-Mail: prabakaran.s@ktr.srmuniv.ac.in

**ABSTRACT**

Analytics often involves studying past historical data to research potential trends. Weather condition is the state of atmosphere at a given time in terms of weather variables like rainfall, cloud conditions, temperature, etc., the existing models use data mining techniques to predict the rainfall. The main disadvantage of these systems is that it doesn't provide an estimate of the predicted rainfall. The system calculates average of values and understand the state of atmosphere, which doesn't yield estimate results. This paper represents a mathematical method called Linear Regression to predict the rainfall in various districts in southern states of India. The Linear Regression method is modified in order to obtain the most optimum error percentage by iterating and adding some percentage of error to the input values. This method provides an estimate of rainfall using different atmospheric parameters like average temperature and cloud cover to predict the rainfall. The linear regression is applied on the set of data and the coefficients are used to predict the rainfall based on the corresponding values of the parameters. The main advantage of this model is that this model estimates the rainfall based on the previous correlation between the different atmospheric parameters. Thus, an estimate value of what the rainfall could be at a given time period and place can be found easily.

**Keywords**: rainfall prediction, linear regression.

## 1. INTRODUCTION

The application of science and technology that predicts the state of atmosphere at any given particular time period is known as Weather forecasting. There is a many different methods to weather forecast. Weather forecast notices are important because they can be used to prevent destruction of life and environment. The weather forecasting methods used in the ancient time usually implied pattern recognition i.e., they usually rely on observing patterns of events. For example, it is found that the following day has brought fair weather; if the preceding day sunset is particularly red. However, all of the predictions prove not to be reliable. Here in this system we used parameters like average temperature, cloud cover to predict the rainfall. The data set of 100 years is taken for this project and the implemented using Numerical methods. Weather forecasting system uses atmospheric parameters like humidity, wind and temperature and forecast weather based on preceding record, thus, this forecasting is more reliable. There are many application that this system be used such as Air Traffic, Agriculture, Marine, Forestry, Navy, and Military etc. The training of the data is done using a modified version of linear regression method. The error percentage between the actual and predicted is used to improve the training set and train the data with the new inputs Thus, with the advantages this method, we can forecast the outcomes till there is no further improvement in the error percentage. This method can be used to forecast the rainfall and prevent the destruction caused by it to the life or property.

## 2. BACKGROUND

### a) 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

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where y denotes the dependent variable (rainfall) and $x_i$ 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

$$\text{Rainfall} = (\text{AvgTemp}*\beta_1) + (\text{CloudCover}*\beta_2) + \beta_3$$

where $\beta_1$ , $\beta_2$ , $\beta_3$ represents the different coefficients for different districts.

### b) 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] developed a forecasting model using mathematical regression. The weather data is collected for a period of 3 years and this model can predict max and min temperatures for a period of 15 to 45 weeks into the future.

Goutami [5] used Multiple Linear Regression to estimate average summer – monsoon rainfall on the data from 1871 to 1999. She analysed the monthly rainfall of Indian summer monsoon months.

Kannan, Ramachandran and Prabhakaran [6] implemented Multiple linear regression and Karl Pearson coefficient. They made a short – term forecast over a particular state. They used fuzzy sets, neural networks to analyse the data.

Retius and Delson [7] developed a weighted multiple regression model. They used combination of time series analysis and regression to offer a powerful system for predicting annual rainfall.

Timothy and Shukla [8] proposed the F-test and Screening procedure. a cross-validation procedure is used first to screens models out that are all likely to poorlyperform on independent datasets, then the error of each model is compared with those all other models to determine threshold of significance in error variance.

Guhathakurta [11] used dynamic models on nonlinear equations that atmospheric system governs. They implemented neural networks with three layer that works on one input, one output and one hidden layer. The network training is carried out till mean square error of 0.0005 to 0.001.

Nikhil [12] proposed correlation and regression both linear and multiple linear regression. He estimated the rainfall by analyzing the atmospheric factors like precipitation, vapour pressure, average temperature and cloud cover.

**c) 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. Add 20 percent of error percentage to the input training rainfall value and now these values are used in the linear regression method to train i.e., steps 3 and 4 is executed.
   b. If the results obtained are not satisfied, we can repeat the step 3, 4, 5 till the point where by adding the error percentage there will be no improvement in the predicted rainfall.
   c. We can try the above step by increasing the error percentage value that is added to the training set, i.e., say we are not satisfied with the error percentage after the first iteration, so we can increase the error percentage that is added in the next iteration to 40 % and increase so on..
   d. In the above steps the coefficients of the independent variables in the linear regression keep on changing.

6. The latest updated coefficients to forecast the test data and this produces the most accurate forecast values.

**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 |

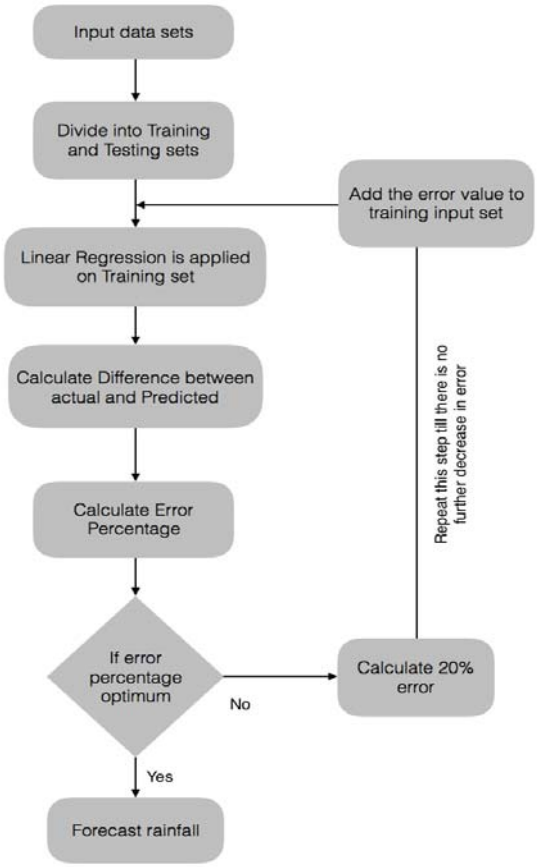


**Figure-1.** System architecture.

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 |

In order to obtain the average temperature and cloud cover value of future time period the linear regression is applied to average temperature and year to obtain the equation to calculate the average temperature, similarly for the cloud cover.

The following results are obtained after the fifth iteration

**Table-3.** Forecast result.

| Year | Actual Rainfall | 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. CONCLUSIONS

Rainfall is the major cause for many of the natural disasters like flash floods, droughts, tsunamis. So in order to prevent these natural calamities, we should be able to predict the cause of the source. The proposed


system can be used to estimate the rainfall over the required period so that the respective authorities can take precautions to prevent the loss of life and property. The proposed system uses modified linear regression approach to predict the rainfall that has less error percentage than compared to most data mining techniques like clustering, back propagation which provides the generalized values rather than estimate values. This data is used to perform the necessary calculations to predict the rainfall from average temperature and cloud cover of that particular district. We may improve this system further using multiple regression which can take multiple months at a time as input and just forming a single equation which leads nearer to an accurate rainfall predicted. The proposed approach may also be used in other applications like, in schools to predict the average marks of their students, in sports to predict the scores or winning teams based on their previous performance, in enterprises to estimate their profits, etc.,.

**REFERENCES**

[1] Daniela Şchiopu, Elia Georgiana Petre, Catalina Negoina "Weather Forecast using SPSS Statistical Methods" This paper presents a case study of using SPSS 13.0 in weather prediction. Vol. LXI No. 1/2009 Gas University of Ploiesti, Bd. Bucuresti, No. 39, 100680, Romania.

[2] R. Samuel Selvaraj and Raajalakshmi "Statistical Method of Predicting the Northeast Rainfall of Tamil Nadu", Universal Journal of Environmental Research and Technology. Volume 1, Issue 4: 557-559.

[3] Dhawal Hirani and Dr. Nitin Mishra "A Survey on Rainfall Prediction Techniques" International Journal of Computer Application (2250-1797) Volume 6– No.2, March- April 2016.

[4] Paras and Sanjay Mathur "A Simple Weather Forecasting Model Using Mathematical Regression" 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, P. Ramchandran "Rainfall Forecasting Using DataMining Techniques". International Journal of Engineering and Technology Vol.2 (6), 2010, 397-401.

[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".

[9] India Metrological department (1973): Northeast monsoon; FMU Rep.no .iv:18.4.

[10] S. Balachandran., R. Asokan, and S. Sridharan, (2006): Global surface temperature in relation to Northeast monsoon rainfall over Tamil Nadu. J. Earth. syst. Sc., vol 115: 349-362.

[11] Guhathakurta, P (2005) "Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network", Current Science, 90, 773-779.

[12] Nikhil Sethi *et al*, " Exploiting Data Mining Technique for Rainfall Prediction" in International Journal of Computer Science and Information Technologies ISSN:0975 9646 Vol. 5 (3) , pp. 3982-3984, 2014.

[13] Guhathakurta, P (2005) "Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network", Current Science, 90,773-779.

[14] Hastenrath, S (1988) "Prediction of Indian Monsoon Rainfall: Further Exploration", Journal of Climate, 1, 298-304.

[15] Montgomery, D. C. and Lynwood, A. J., 1996, Forecasting and Time Series Analysis, New York, McGraw-Hil.