# Rainfall Prediction using Linear Regression

**Amruta Prabhakar Dhanfule[1] Bhakti Dadasaheb Thombare[2] Ritu Vinodbhai Kumbhani[3]**
[1,2,3]BE Student
[1,2,3]Department of Information Technology
[1,2,3]DYPCOE, India

*Abstract*— This project is carried on the heuristic prediction of rainfall using machine learning techniques. As we all know agriculture was the leading issue of our country and economy. Even though a regular rain pattern is usually played important role for healthy agriculture but too much rainfall or too little rainfall can be injurious, even it led to destructive for crops. The decreasing trends in seasonal rainfall and post-monsoon rainfall and increasing occurrence of the deficit rainfall years indicates the probable intensification of water scarcity. This project calculates the rate of rainfall in previous years according to various crops season like rabbi, Kharif, zaid and predicts the rainfall in future seasons. We have selected a real data set which consists of past year's rainfall rate accorded to various seasons. This website results in helping farmers to make a correct decision to harvest a particular crop accordingly to crops seasons.

***Key words:*** Rainfall Prediction, Linear Regression

## I. INTRODUCTION

India's foremost occupation is agriculture and its economy depends upon the agriculture of the country. But rainfall has a dramatic effect on agriculture. So, to make rainfall prediction earlier is important for the better growth of the economy. Early prediction of rainfall has been one of the most competitive tasks in the world from previous years. Various kinds of techniques used for prediction is Artificial Neural Network (ANN), Support Vector Machine (SVM), Regression analysis, and clustering etc. But the crop seasons in India, Pakistan and Bangladesh were distributed them into three main seasons as rabbi, Kharif, and zaid or zayad. The terms are derived from Arabic where rabbi means spring, Kharif means autumn and zaid means summer. Periods of this season is Kharif from July to October during the time of southwest monsoon, a rabbi from October to march and zaid from march to July it's mainly during summer. Every season has its own harvesting crops were in rabbi contains wheat, oats, onion, tomato, potato, peas, barley, linseed, mustard oil seeds and masoor etc., Kharif contains rice, sorghum, groundnut, jowar, soya bean, bajra, jute, maize, cotton, hemp, tobacco ragimillet, and arhar etc. and zaid contains sugar cane, cucumber, rapeseed, sunflower, rice, cotton, oil seeds, watermelon, and muskmelon etc. Linear regression is a technique for predicting the various value of a dependent variable from an independent variable when the relationship between the variables can be described by a linear model. It is a method of estimating the conditional expected value, for example, one crop season rabbi given the values of some other seasons like Kharif. The variable of interest rabbi is conventionally called the dependent variable.

## II. LITERATURE SURVEY

*A.* Title: Introduction to Rainfall Prediction Using Data Mining Technique

1) Review

A wide variety of rainfall forecast methods are available in India, because India is an agricultural country and the success of agriculture depends on rainfall and humidity. There are mainly two approaches to predict rainfall in India. They are Empirical method and dynamical method. The empirical approach is based on analysis of historical data available from various government sites of the rainfall and its relationship to a variety of atmospheric variables. The most widely used empirical approaches used for climate prediction is regression, Artificial Neural Network(ANN), Decision Tree algorithm, fuzzy logic and group method of data handling. In dynamical approach forecasting is done by physical models based on systems of equations. These equations predict the assessment of the global climate system in response to initial atmospheric conditions. There were various algorithms have been analyzed. Data mining techniques are efficiently used in rainfall prediction. Data mining is a technique which finds useful patterns from huge amount of data. Data mining can also be defined as the process of extracting implicit, previously unknown and useful information and knowledge from large quantities of noisy, ambiguous, random, incomplete data for practical application. It is a powerful new technique with high potential to help companies focus on the most important information in their databases. It uses machine learning, statistical and visualization technique to discover and predict knowledge which is understandable to the user. Prediction is the most important technique of data mining which discovers relationship between exploratory and response data.

Weather forecasting is the application of science and technology to predict the rainfall for a given location. It is becoming very important for scientists, agriculturists, farmers, global food security, disaster management and related organizations to understand the natural phenomena to plan and be prepared for the future. The weather prediction began with early civilizations using reoccurring astronomical and meteorological events to help them monitor seasonal changes in the weather. From very recent time, attempts have been made to produce forecasts based on weather changes and personal observations. Title: Support Vector Machine for Rainfall Prediction Review: The SVR model maps data non-linearly into a higher dimensional feature space, in which it undertakes linear regression. Rather than obtaining empirical errors, SVR aims to minimize the upper limit of the generalization error. Suppose we are given training data (x$i$, $i$) $i$=1, where $xi \in Rn$ is the input vector; $di$ is the output value and $n$ is the total number of data dimension. The modeling aim is to identify a regression function, $y = (x)$, that

 

accurately predicts the outputs $di$ corresponding to a new set of input– output examples, ($xi$, ). The linear regression function is described as follows:

$(x)=\omega(x)+b,\ i:\ Rn \rightarrow F,\ \omega \in F$ ……(1)

where$\omega$ and $b$ are coefficients; $\phi(x)$ denotes the high dimensional feature space, which is nonlinearly mapped from the input space x. This primal optimization problem is a linearly constrained quadratic programming problem, which can be solved by introducing Lagrangian multipliers and applying Karush-Kuhn-Tucker (KKT) conditions to solve its dual problem $minR(\alpha,\ \alpha)= N \sum i=1\ (\alpha i -\alpha^* i)-\varepsilon N \sum$ i=1 $di(\alpha i + \alpha^* i)-1/\ 2\ N \sum i=1\ N \sum j=1\ (\alpha i -\alpha*)(\alpha j -\alpha* j)K(xi,xj)\ s.t.\ N \sum i=1\ (\alpha i -\alpha*)= 0,\ 0 \le \alpha,^* \ge C, i =1, 2,\cdots\cdots,$N…… (2) where $\alpha i$ and $\alpha* i$ are the Lagrangian multipliers associated with the constraints, the term $K(xi,xj)$is defined as kernel function, where the value of kernel function equals the inner product of two vectors $xi$ and $xj$ in the feature space $\phi(xi)$ and $\phi(xj)$, meaning that $K(xi,xj))=\phi(xi) \times \phi(xj)$. In machine learning, the popular kernel functions are linear kernel, Polynomial kernel and Guassian kernel. Title: Rainfall Prediction Using Regression Approach Review: The application of science and technology that predicts the state of atmosphere at any given particular time period is known as Weather forecasting. There are a many methods to weather forecast. Rainfall prediction is very important because they can be used to prevent destruction of life and environment. The rainfall prediction methods used in the recent 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 clear weather; if the preceding day sunset is particularly yellow. However, all the predictions prove not to be reliable. Here in this system we used parameters like average temperature, humidity to predict the rainfall. The data set of 50 years is taken for this project and the implemented using Numerical methods. Rainfall prediction system uses a atmospheric parameter like humidity, wind and temperature and predict rainfall based on preceding record, thus, this prediction is more reliable. There are many applications that this system be used such as Air Traffic, Agriculture, Marine, Forestry and crop recommendation etc. The training of the data is done using a modified version of linear regression method which uses multiple exploratory variable. The error between the actual and predicted is used to improve the training set and train the data with the new inputs Thus, with 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.

### *B. Title: Prediction of monthly Rainfall on Homogeneous Monsoon Regions of India*

#### *1) Review*

The aim of the topic is to find the monthly rainfall using homogeneous techniques in India for prediction of rainfall. Introduction it is scientifically and mathematically challenging to use climate signals for the prediction of basin-scale hydrologic variables, because the climatic systems are very complex and physics of many systems is still not very clearly understood. The difficulties in modeling such complex systems are considerably reduced by the recent Artificial Intelligence tools like Artificial Neural Networks (ANNs); Genetic Algorithm (GA) based evolutionary optimizer and Genetic Programming (GP). Hence such AI tools are tried nowadays for modeling complex systems like basin-scale stream-flow forecasting using the information of large-scale atmospheric circulation phenomena. Indian Summer Monsoon Rainfall is always found to vary annually leading to profound impacts on agriculture based Indian economy. Generally meteorological forecasts are generated for three timescales, viz. short-range (1–2 days ahead), medium-range (3– 10 days ahead) and long-range forecasts for monthly and seasonal scales. In India, India Meteorological Department (IMD) generates the short and long-range predictions, whereas the National Centre for Medium Range Weather Forecasting (NCMRWF), New Delhi is responsible for the medium-range predictions. Prediction of ISMR is having a long history. It started with the work of Sir Henry Blanford in 1886, which was entirely based on Himalayan snowfall. John Eliot used extra-Indian factors, viz. Pressure of Mauritus, Zanzibar and Seychelles in the monsoon forecast of 1896. Sir Gilbert Walker proposed statistical association for monsoon forecast. He systematically examined the relationship between Indian monsoon rainfall and global circulation parameters. He selected 28 predictors to issue forecast based on regression equation during the year 1906 (Jagannathan, 1960; Rao and Rama Moorthy, 1960; Rao, 1965). Most of the Walker's predictors (except Himalayan snow accumulation) were signs of different facets of Southern Oscillation (Shukla and Paolino, 1983). Savur (1931) showed that 7 out of the 28 parameters had lost their significance in due course of time. Since then, extensive research work has been carried out on empirical seasonal forecasting of Indian Summer Monsoon Rainfall. Some of the noteworthy studies can be listed as following: Banerjee et al. (1978), Kung and Sharif (1982), Bhalme et al. (1986), Gowariker et al. (1989, 1991), Parthasarathy et al. (1988, 1991, 1995), Krishna Kumar et al. (1995, 1997), Rajeevan et al. (2004), etc. In spite of many efforts in the long range prediction of all-India summer monsoon rainfall (AISMR), it is felt that achieved success is not adequate and there is much scope to investigate new predictors and new methodologies of ISMR prediction.

## III. FUTURE SCOPE

The application of science and technology that predicts the state of the atmosphere at any given particular time period is known as Weather forecasting. There are many different methods to a weather forecast. Weather forecast notices are important because they can be used to prevent the destruction of life and environment. So basically this project helps farmers to decide the proper crops for their fields. For that problem, we use a machine learning technique to first determine the rainfall and after that farmers can get the guidance for their crops.

### *A. Proposed System*

Agriculture is the predominant of our country and economy. While a regular rain pattern is usually played very important role for healthy agriculture but too much rainfall or too little rainfall can be harmful, even it led to devastating of crops. So

 

we analyze the rate of rainfall in previous years according to various crops seasons like Rabi, Kharif, zaid and predicts the rainfall in future seasons.

The aim of the project is predicting rain before the monsoon comes so that the farmers can choose their crops according to the rainfall prediction.

### *B. System Specification*

This section demonstrates the Rainfall Prediction Model's most prominent features and explains how they can be used and the results they will give back to the user.

*1) Hardware Requirements*

The minimum hardware requirements of Rainfall prediction are a 500 Megahertz CPU and 128megabytes of RAM. Also, because this model uses python libraries to speed up the prediction. A system with a large amount of memory for training and testing data.

*2) Software Requirements*

- Windows 2000
- Windows XP
- Windows Vista
- Windows 7
- Windows 8
- Windows 10
- Mac OS X
- Linux

### *C. System Architecture*

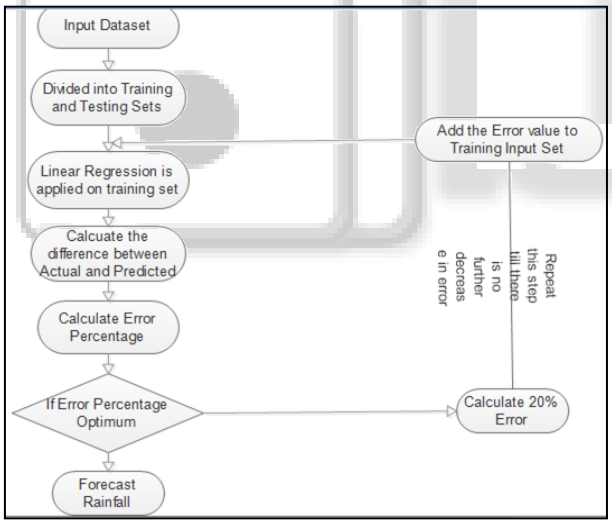


Fig. 1: System Architecture of Proposed System

## IV. PROPOSED SYSTEM

### *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

Y1=B01+B1 xa+……+B p xip+ci

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

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

Rainfall = (AvgTemp*B1) + (CloudCover*B2) + B3

Where B1, B2p, B3represents the different coefficients for different districts.

### *B. Data Collection*

The data required for the system to predict rainfall is previous year rainfall data, average temperature and humidity over of the particular area. The data is collected for a period of 15 years i.e., from 2000 to 2015 for each month. This data is then used 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 are required to predict the rainfall. This data is used to formulate the equation for predicting the rainfall by calculating the average temperature and humidity at that particular area. 5.3 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 is examined. The input data of training set is obtained from 2000 to 2015 for each month to perform the proposed system and check the method.
2) The training and test data is 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 2016 to 2018 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 performed are given by

a) Add 20 percent of error percentage to the input training rainfall value and now these values are used in the multiple 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 multi linear regression keep on changing.

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

*C. Server*

- Obtain information
- Display information on user's app
- Update information according to user instructions

*1) Obtain Sensor Information*

The server takesinformation about previous rainfall.

*2) Display Sensor Information on User's App*

Server sends the data to the user and gives user all the information.

*3) Update information according to user instructions*

The job of server is to use data from database as per instructions sent by user through the website.

*D. User*

- Registration
- Login
- Access to the server

*1) Registration*

User will register to the system with normal information.

*2) Login*

For login the user will enter the user name and password, if entered information is correct then the system will redirect to the home page, otherwise it will show an error message.

*3) Access to the server*

- After login the user will see the detail about temperature
- Checks for the needed information.

## ACKNOWLEDGMENT

The estimation of rainfall is of great importance in terms of water resources management human life and their environment. It can be met with the approximate or incomplete estimation problems because rainfall estimation is affected from the geographical and regional changes and properties. This paper presents review of different methods used for rainfall prediction and problems one might encounter while applying different approaches for rainfall forecasting. Due to nonlinear relationship in rainfall data and ability of learning from the past makes Artificial Neural Network a preferable approach from all available approaches.

## REFERENCES

[1] Ms. AshwiniMandale, Mrs.Jadhawar B.A, "Weather Forecast Prediction: A Data Mining Application", International Journalof Engineering Research and General Science Volume 3, Issue 2, March- April, 2015, ISSN 2091-2730.

[2] Nikhil Sethi, Dr.KanwalGarg,"Exploiting Data Mining Technique For Rainfall Prediction", Nikhil Sethi Et Al, / (IJCSIT) International Journal Of Computer Science And Information Technologies, Vol. 5 (3), 2014, 3982-3984.

[3] Duncan Waga And KefaRabah, "Environmental Conditions' Big Data Management And Cloud Computing Analytics For Sustainable Agriculture", PrimeResearch On Education (Pre) ISSN:2251-1251, Vol. 3(8), Pp. 605-614, November 30th, 2013.

[4] T V RajiniKanth, V VSssBalaramAnd N.Rajasekhar,"Analysis Of Indian Weather Data Sets Using Data Mining Techniques", Pp. 89–94, 2014.© Cs & It-Cscp 2014.

[5] Mohini P. Darji, Vipul K. Dabhi and HarshadkumarB.Prajapati, Rainfall Forecasting Using Neural Network: A Survey, ICACEA-2015, Ghaziabad, India.

[6] Elia Georgiana Petre, "A Decision Tree for Weather Prediction", Vol. LXI, No. 1/2009, 2009, pp. 77-82.

[7] C. Venkatesan, S.D. Raskar, S.S. Tambe, B.D. Kulkarni, R.N. Keshavamurty, 1997, Prediction of all India summer monsoon rainfall using error-back-propagation neural networks. Meteorol. Atmos. Phys. 62 (3–4), 225–240.

[8] P. Goswami and Srividya, "A novel Neural Network design for long range prediction of rainfall pattern," Current Sci.(Bangalore), vol. 70, no. 6, pp. 447-457, 1996.

[9] S. Kannan , SubimalGhosh, "Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output", Springer-Verlag, July- 2010.

[10] M. A. Sharma and J. B. Singh, "Comparative Study of rainfall orecasting models," New York Sci. J., pp. 115-120, 2011.

 