Research article Available online www.ijsrr.org ISSN: 2279–0543

International Journal of Scientific Research and Reviews

A Machine Learning Approach for Annual Rainfall Prediction Using Linear Regression Model

Sushuma Narkedamilli[1*], Karri L Ganapathi Reddy[2] and V H L Manikyambha[3]

Department of Computer Science & Engineering, BVCITS, Batlapalem-533221, A.P., INDIA.
E-mail: chakkasushuma@gmail.com, Mobile: 9290541330

ABSTRACT

Machine learning is a branch of Artificial Intelligence (AI) that provides systems to learn automatically without human intervention or assistance and adjust actions accordingly. This learning process begins with observations or data, such as direct experience, or instruction, in order to look for patterns in data and make better decisions in the future. This has motivated the research field of machine learning that is concerned with observations or data learning methods. Here we have taken a problem regarding Annual Rainfall prediction. In our paper, we represent a method named Linear Regression to predict the annual rainfall in various states in India[4]. To predict the estimate of annual rainfall, here we consider atmospheric parameter, average temperature. The linear regression is applied on the data set and the coefficients are used to predict the annual rainfall based on the corresponding parameter values. Finally an estimate value of what the rainfall could be at a given values and place can be found easily.

KEYWORDS: Linear Regression, Learning Process, Machine Learning

***Corresponding Author:**

Mrs. Sushuma Narkedamilli

Department of Computer Science & Engineering,

Bonam Venkata Chalamayya Institute of Technology & Science,

Batlapalem, A.P., INDIA.

Email: chakkasushuma@gmail.com, Mob No-9290541330

INTRODUCTION

Rainfall play important role in forming of a balance in natural life. It plays a significant role in agriculture and farming and undoubtedly; water is one of natural resources on earth. The changing climatic conditions and the increasing greenhouse emissions have made it difficult for the human beings and the planet earth to experience the necessary amount of rainfall that is required to satisfy the human needs. Rainfall prediction have became an important issue and challenge. An analysis of rainfall data for the past five years, available on the website of the India Meteorological Department, points to changes in rainfall trends in India. The trends in rainfall do not indicate the kind of disastrous rainfall received this year, but it does point to the changing necessity for a robust disaster management program, which as of now does not exist.. Therefore, to analyze the changing patterns of the rainfall, we try to predict the rain not only for the human needs but also for natural disasters that could cause by the unexpected heavy rainfalls. To be more specific and aware of the drastic and destructive climatic changing predicting rainfall has been the focus of computer scientist and engineers. This paper focuses on predicting annual rainfall using Linear regression.

The annual rainfall predictions not only analyze the changing patterns of rainfall but also help in taking the precautionary measures in case of disaster and its management. Similarly, the annual rainfall prediction and weather updates not only help in managing the macro level problems like flood and agricultural issues because of poor or heavy rainfall. The rainfall prediction could also contribute to the well-being and comfort of the people by keeping them informed by tracking the rainfall patterns and predicting the rainfall using Machine Learning. The annual rainfall predictions help the people to deal with different weather conditions like hot and humid weather. The Technological development in the modern world has expanded the space for innovation and revolution.

LITERATURE REVIEW

1. In this article, we make a comparison of several machine learning methods of forecasting an average daily and monthly rainfall in Japan. All the methods are coupled with two data-preprocessing techniques[6]. Prior to applying the methods, two input selection techniques are used. For the modeling of the rainfall, a novel hybrid multi-model method is proposed[5]. The hybrid method generates sub-models first from each of the above methods with different parameter settings. Second, all the sub-models are ranked with a variable selection technique called least angle regression (LARS). Third, the higher ranked models are selected based on their Leave-One-Out Cross-Validation (LOOCV) error. The forecasting using the out of samples is done by a weighted combination[1] of the finally selected models. The constituent models of the hybrid method are the

ANN, Multivariate Adaptive Regression Splines (MARS)[4], the *k*-nearest neighbor, and radial basis Support Vector Regression (SVR) [2]. For evaluation of this hybrid method, we have constructed all these methods with their respective optimal parameters and applied to out of sample forecasting. Here we discuss briefly the study area and the rainfall series used in this paper[3]. We describe the hybrid forecast model including the input selection technique and the variable selection method, and how the weights are extracted. This is followed by discussions about the experimental setup and results.

3. PROPOSED SYSTEM

Our proposed "Annual Rainfall Prediction" represents a mathematical method called Linear Regression to predict the rainfall in various states in 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 annual 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. Thus, an estimate value of the annual rainfall at a given place can be predicted accurately.

Advantages :

i. Aircraft and shipping relay heavily on accurate prediction
ii. Surfers can know when large waves are expected.
iii. Farmers can know when to plant or harvest their crops.
iv. Regions evacuated if hurricanes or floods are expected.

SYSTEM ARCHITECTURE:

Data collection:

The dataset we used in this paper was from KaggleInc which is an open source dataset and consists of 4000 records with 19 parameters. However out of these 19 parameters only 4 were chosen which are bound to affect the annual rainfall. Parameters such as Jan-Feb, Mar-May, are independent variables. Annual is a dependent variable on several other independent variables.

Preprocessing:

It is a technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends is likely to contain many errors. It involves the process of finding out missing and redundant data in the dataset.

Here Entire dataset is checked for NaN. The observations NaN will be deleted. Thus, this brings uniformity in the dataset. However in our dataset, there was no missing values found meaning that every record was constituted its corresponding feature values.

Data classification:

Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.

Data regression:

Regression is basically a statistical approach to find the relationship between variables. In machine learning, this is used to predict the outcome of an event based on the relationship between variables obtained from the data-set. Linear regression is one type regression used in Machine Learning.

Prediction of Output:

Output can be predicted by using Machine Learning algorithms.

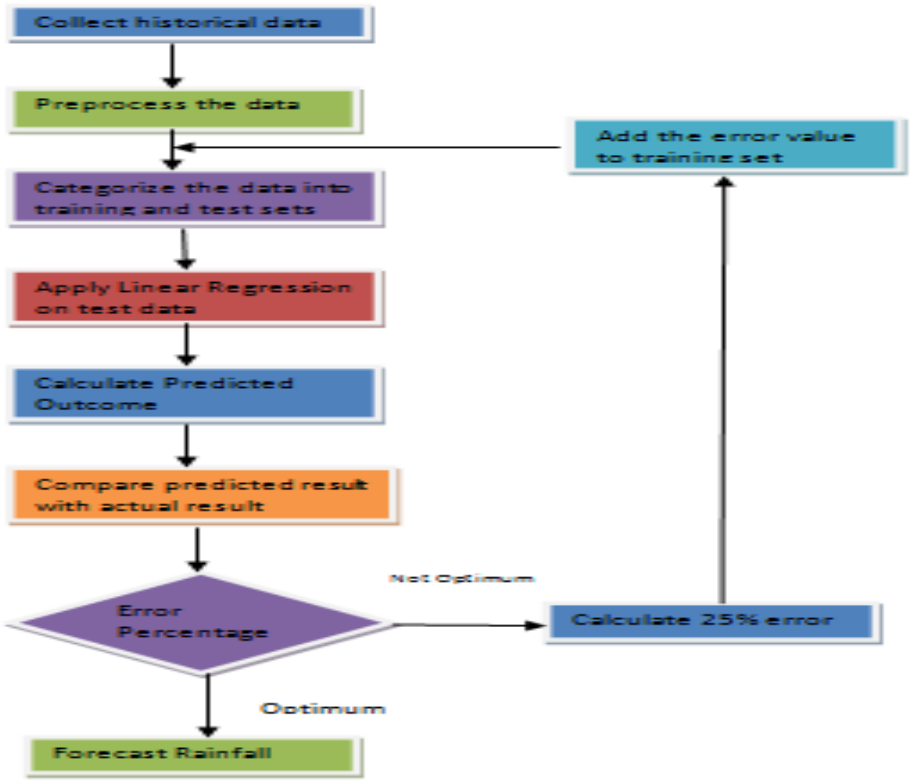


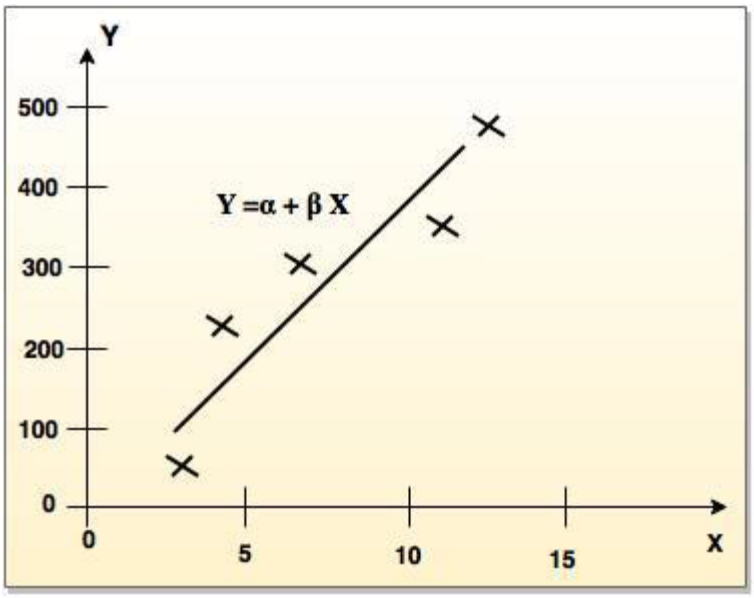
Figure1. System architecture

LINEAR REGRESSION

Linear Regression is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.

Hypothesis function for Linear Regression : $y= \alpha + \beta.x$



Linear Regression

While training the model we are given:

x: input training data (univariate – one input variable(parameter))

y: labels to data (supervised learning)

When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best **α** and **β** values.

α: intercept

β: coefficient of x

Once we find the best **α** and **β** values, we get the best fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x.

How to update α and β values to get the best fit line?

Cost Function (J):

By achieving the best-fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the **α** and **β** values, to reach the best value that minimize the error between predicted y value (pred) and true y value (y).

$$minimize \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

Cost function(J) of Linear Regression is the **Root Mean Squared Error (RMSE)** between predicted y value (pred) and true y value (y).

Gradient Descent:

To update **α** and **β** values in order to reduce Cost function (minimizing RMSE value) and achieving the best fit line the model uses Gradient Descent. The idea is to start with random **α** and **β** values and then iteratively updating the values, reaching minimum cost.

Algorithm for Linear |Regression:

```
SUB Regress(x, y, n, a1, a0, syx, r2)
sumx = 0: sumxy = 0: st= 0
sumy = 0: sumx2 =0: sr = 0
DO i = 1, n
sumx = sumx + xi
sumy = sumy + yi
sumxy = sumxy + xi*yi
sumx2 = sumx2 + xi*xi
END DO
xm = sumx/n
ym = sumy/n
a1 = (n*sumxy – sumx*sumy)/(n*sumx2 – sumx*sumx)
a0 = ym – a1*xm
DO i = 1, n
```

st = st + (yi – ym)^2

sr = sr + (yi – a1* xi – a0)^2

END DO

syx = (sr/(n – 2))^0.5

r2 = (st – sr)/st

END Regress

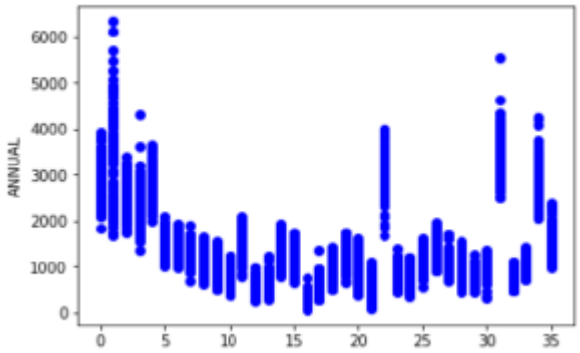


Figure2. Scatter Plot for subdivision and annual rainfall

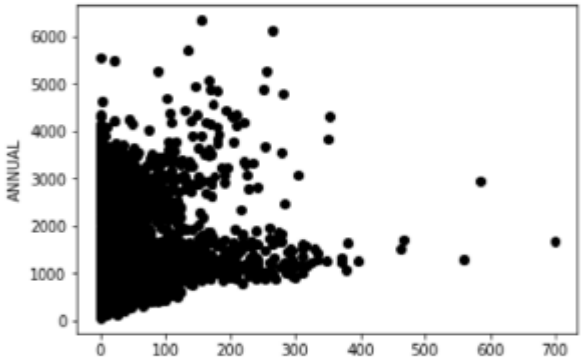


Figure3. Scatter Plot for Jan-Feb and annual rainfall

6. CONCLUSION

Rainfall is one the most significant natural phenomenon that is not only important for the human beings but also for the all other living beings. Our study aimed at building a predicting system using linear regression that could predict annual rainfall accurately and efficiently with minimum error. Finally using linear regression algorithms we analyzed and predicted the annual rainfall.

REFERENCES

1. Timmermann, A. *Forecast Combinations, Handbook of Economic Forecasting*. Elsevier; 2006; 1(4): 135–196.
2. Cherkassky, V. and Ma, Y., "*Practical selection of SVM parameters and noise estimation for SVM regression*", Neural Networks: The Official Journal of the International Neural Network Society, 2004; **17**(1): 113–26.
3. S. Monira Sumi , M. Faisal Zaman,Hideo Hirose**. "***A Rainfall Forecasting Method Using Machine Learning Models And Its Application To The Fukuoka City Case"*, Int. J. Appl. Math. Comput. Sci., 2012; 22(4): 841–854
4. Myers, R.H. "*Classical and Modern Regression with Applications"*, Duxbury, Boston, MA. 1990

5. Chalimourda, A., Sch¨olkopf, B. and Smola, A.J, "*Experimentally optimal v in support vector regression for different noise models and parameter settings*", Neural Networks:The Official Journal of the International Neural Network Society, 2004; **17**(1): 127–41.
6. Shabib Aftab, Munir Ahmad, Noureen Hameed, Muhammad Salman Bashir, Iftikhar Ali, Zahid Nawaz, "*Rainfall Prediction using Data Mining Techniques: A Systematic Literature Review*", International Journal of Advanced Computer Science and Applications.
7. R. Kohavi and F. Provost, "Glossary of terms," Machine Learning, 1998; 30(2): 271–274.