IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 2, February 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102020 |

Rainfall Prediction using Machine Learning Techniques

M.C.S.Geetha, Kirsten Kennedy

Assistant Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, India

PG Student, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, India

ABSTRACT: Rainfall is important for human society, so predicting the rainfall becomes the challenging tasks to keep the people healthy. Timely and accurate forecasting is essential to avoid financial loss and to minimize difficulties of drought and flood. Several research has been done using algorithms of machine learning for various datasets. In this paper, various machine learning methods like Extreme Gradient Boost, Random Forest and Multivariate Linear Regression were used to analyse the performance and to decide which is better. The dataset was collected from Kaggle website. The performance was evaluated by using the methods such as Mean absolute Error and Root mean squared error. It showed that, the Extreme Gradient Boosting algorithm is better compared to Random Forest and linear Regression, based on the outcome of the study.

KEYWORDS: Machine Learning, Forecasting Rainfall, Multivariate Linear Regression, Extreme Gradient Boost, Random Forest

I. INTRODUCTION

Rainfall forecasting is very essential factor, or else lot of calamities may occur. If there is unusual rainfall, there is damage causes in the crops. So, there is need for accurately calculate the rainfall in order to effectively use the water sources, crop production, and also planning water structures in advance. The important source of India is Agriculture. The technology in the field of agriculture have been improved over the last 10 years, and there is a growth in the rate of worldwide warming, air pollution, noise, water, dust, etc. This caused the climate and weather conditions to have extreme variations. Rainfall plays a major role for harvesting crops. If there is change in the pattern, if will automatically affects the water supply. Thus, the research on the fluctuations in the rainfall rates is the most natural water source administration. The Machine Learning algorithms is mostly used for predictions. With the help of these algorithms, the forecasting can be done accurately. The correct algorithm selection must be done based on the problem statement. Regression techniques in the Machine Learning can be used for prediction of all different kinds of datasets. Regression analysis can be used for correctly analyse the data and to provide with exact prediction of the weather and rainfall. It can also be used for prediction in different fields in business, marketing and finance.

In the daily life, Rainfall plays an important factor. The occurrence of flood leads to damage in the infrastructure and in the human life. The social and economic factors have greater impact on the rainfall [1]. Currently, forecasting of the Rainfall have become the very vital component for best water storage schemes present in the worldwide. The ambiguity of the rainfall information leads to many difficult challenges. Nowadays, generally rainfall prediction techniques could not identify the unknown models present in the rainfall information. This paper helps to find out every unknown models and that would help for forecasting the rainfall accurately [2].

Many difficulties to find out the unknown models was there in the existing techniques. So, the forecasting was wrong and leads to heavy losses. Hence, this study helps to forecast the rainfall effectively by accurate predictions and solve all the problems and thereby economy is developed with the agriculture [3].

This research paper is structured as follows. First, we explain the techniques of the machine learning in Section 2. The literature survey of the work are presented in Section 3, while the experiments and results are shown and discussed in Section 4. At the last, conclusions are given.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102020 |

II. MACHINE LEARNING METHODS

2.1 Random Forest Regression Model

This model is a supervised technique and it helps to predict all kinds of problem in correct manner. It also includes the characteristics of non-linear relationship based on the selection of the problem. It uses collective learning method for prediction. It trains the algorithm by making decision trees and thereby result is obtained by forming the mean of the classes together with trees. The decision trees are constructed by first taking a random data from the training set and building tree by connecting the data points. More trees can be drawn by repeating again the same process. This decision trees helps to make a decision and find out the prediction of the rainfall.

2.2 XGBoost gradient descent

This extreme gradient boosting uses the supervised machine learning problem having various features of xi and forecasts a target variable yi. This algorithm is used because of its speed and accuracy for various regression and also classification problems for finding the forecast. This algorithm forms the best and efficient for which has linear model as well as tree learning algorithm. It is quicker since it can do similar calculation in one machine. Forecasting the rainfall can be done quickly with the help of these algorithms. This algorithm learns by the Distributed and Parallel computing and can be made efficient which provide robust solution.

2.3 Multivariate Linear Regression

In this regression, there are many independent variables and only one dependent variable. The result is forecasted based on the number of independent variables. In this paper, it contains many independent variables and it is applied to predict the rainfall. This regression uses the machine learning techniques to forecast the rainfall in the area. By analysing using this algorithm, we found how effectively it predicts the strength of the rainfall.

The equation is given as

$$\mathbf{y = \beta 0 + \beta 1.x1 + \beta 2.x2 + \ldots.. + \beta n.xn}$$

The equation for Rainfall prediction is as follows

$$(\beta 1*year) + (\beta 2*month) + (\beta 3*day) + (\beta 4*MaxTemp) + (\beta 5*MinTemp) + (\beta 6*Humidity) + (\beta 7*Evaporation) + (\beta 8*sunshine) + (\beta 9*windspeed) + \varepsilon i$$

III. RELATED WORK

Support Vector Machine, Linear regression with Neural Network and Random Forest algorithms were used for analysing the rainfall forecasting and found Linear regression gives higher result when compared to other algorithm [4].

The non-parametric approaches for analysing the pattern of rainfall were completely analysed such as Pettitt Test, Method for change point detection,Method for analysing rainfall changes, Mann-Kendal test, rank-based test, Method for innovative trend analysis. The India dataset of the weather conditions was analysed in this paper [5].

A relative analysis was conducted for forecasting rainfall using machine learning algorithms in the area of Nepal. The techniques used were Decision Tree, Support Vector Machine and Random Forest. The accuracy, F-measure, recall and precision was calculated. The Random forest has high accuracy of 80% [6].

[7] The KNN, Support Vector regression and Support Vector Machine were used for predicting the rainfall. The SVM provides better result compared to other techniques. This paper has given the ecological highlights which provides constructive and destructive influence of the rainfall. The amount of rainfall the occurs every day is analysed.

Artificial Neural Network and Emotional Neural Network have been used in the area of Bihar for demonstrating overflow of the rainfall because of flood. The greatest outcome is achieved by the Emotional Neural Network by calculating the correlation[8].

The methods need to be demonstrated by using the Machine Learning techniques for making the improvement of the precision and also the solid value is got by forecasting the technique. The multi-layer perceptron is used for the forecasting the everyday streamflow[9].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102020 |

IV. METHODOLOGY AND FINDINGS

4.1 Methodology

The Machine Learning algorithm was used for forecasting the rainfall. The gradient descent XGBoost, Multivariate Linear Regression and Random Forest algorithms were used for analysing the rainfall accurately and provide the result. After investigating, it is found that the XGBoost provides better result. The performance was considered by using error methods.

The Pearson Correlation when applied, the connection among the variables can be found. If it is zero, there is no relationship between variables. For the value one, there is constructive association present in the variables and for the value one negative, there occurs destructive association among the variables.

The formula is given as

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Where correlation coefficient is represented by r, x-variable values in a sample is represented by xi, x-variablemeanvalues is represented by x', y-variable values in a sample is represented by yi, y-variablemeanvaluesis represented by y'.

The following Table 1 shows the Pearson quantity limits and analyses.

The coefficient of Pearson	Analysis
0.1>0	Insignificant
0.2>0.1	Vulnerable
0.4>0.2	Average
0.6>0.4	Comparatively powerful
0.8>0.6	Powerful
1>0.8	Extremely powerful

Table 1 Pearson quantity limits and analyses

The relevant features by the Pearson correlation coefficient are selected by the machine learning algorithms. The Mean Absolute Error and the Root Mean Squared Error have been used for analysing the rainfall forecasting performance and to find out the algorithm which is better.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

the total number of data points is represented by N, prediction is represented by yi, true valueis represented by xi.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102020 |

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}},$$

The number of data points is represented by N, i-th measurement is represented by y(i) and prediction is represented by y^(i).

4.2 Findings

The environmental features and the pearson correlation coefficient are given in the Table 2. The environmental variables helps to analyse the Pearson correlation. The variables which are linked higher than 0.20 along with rainfall was counted for the environmental experiment of the rainfall forecasting. The environmental attributes outcome of forecasted rainfall such as Sunshine, Evaporation, Maximum Daily Temperature, Humidity, and Minimum Daily Temperature must be taken to forecast the amount of daily rainfall.

From the above Table 2 Pearson Correlation coefficient experimental results, it clearly shows that the day, month, year and wind speed has very less value, so these cannot be considered for the forecasting of the rainfall. So the values above 0.2 are taken for analysing purpose. The greatest value measured from the above table is humidity and sunshine of 0.40 and 0.35. The environmental features were used by the ML model as the input. The regression models are applied in python for analyzing the performances of the algorithm such as XGBoost, RF and MLR and the measurement was done with RMSE and MAE.

Based on the Table 3 above, the performance were analysed for all the above algorithms and found that XGBoost is better. XGBoost algorithms have RMSE value as 7.91 and MAE as 3.60, so this algorithm predicts rainfall well compared to other two algorithms.

Algorithms	Root Mean Squared Error	Mean Absolute Error
XGBoost	7.91	3.60
MLR	8.67	4.99
Random forest	8.90	4.51

Table 3 Performance measurements

V. CONCLUSION

Thus, this paper explains about the forecasting of the rainfall using machine learning techniques since rainfall is one of the essential component for producing crop and in the usage of water resources. The machine learning techniques were used for the predicting the rainfall. The XGBoost, Multilinear Regression and Random forest were used by using the Kaggle dataset. The Pearson coefficient was used with the environmental features for predicting the rainfall. By comparing all the above algorithm, it is found XGBoost provides good result. If the sensor data is used, the accuracy may go high. In this paper, the sensor data is not used for analysis. In future sensor data can be included for the prediction to give more accurate results.

REFERENCES

1. Le T.-T., Pham B.T., Ly H.-B., Shirzadi A., Le L.M.,Development of 48-hour precipitation forecasting model using nonlinear autoregressive neural network, CIGOS 2019, Innovation for Sustainable Infrastructure, Springer (2020), pp. 1191-1196
2. MasafumiGoto, Faizah Cheros, Nuzul Azam Haron, Nur-AdibMaspo, Aizul Nahar Bin, MohdNawiHarun, and MohdNasrun,(2020)," Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review." IOP Science journal.
3. V.P Tharun, Ramya Prakash, S.Renuga Devi.(2019)"Prediction of Rainfall Using Data Mining Techniques" ieeexplore.ieee.org.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 11, Issue 2, February 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1102020 |

4. Mrunmay Jalgaonkar, Dr. Umesh Kulkarni, "Rainfall Prediction using Regressions and Multiple Algorithms", International Research Journal of Computer Science (IRJCS), Volume 8, Issue 4, April 2021.
5. Bushra Praveen, Swapan Talukdar, Shahfahad, Susanta Mahato, Jayanta Mondal, Pritee Sharma, Abu Reza Md.Towfiqul Islam, Atiqur Rahman, "Analyzing Trend and Forecasting of rainfall changes in India using nonparametrical and machine learning approaches", Scientific Report, 2020.
6. Nawaraj Paudel, Tekendra Nath Yogi, "Comparative study of Machine Learning Algorithms for Rainfall Prediction- A case study in Nepal", International Journal of Advanced Research in Engineering and Technology (IJARET), Volume 11, Issue 10, 2020.
7. Arnav G, Kanchipuram Tamil Nadu. Rainfall prediction using machine learning. Int J Innovative Sci Res Technol. 2019. 56–58.
8. S. Kumar, T. Roshni, D. Himayoun, A comparison of Emotional Neural Network (ENN) and Artificial Neural Network (ANN) approach for rainfall-runoff modelling, Civil Engineering, 5 (10) (2019), pp. 2120-2130
9. B. Mohammadi, F. Ahmadi, S. Mehdizadeh, Y. Guan, Q.B. Pham, N.T.T. Linh, *et al.* Developing novel robust models to improve the accuracy of daily streamflow modelling, Water Resource Management (2020)

BIOGRAPHY

Dr.M.C.S.Geetha is an Assistant Professor in the Department of Computer Applications, Kumaraguru College of Technology, Coimbatore. She received Master of Computer Applications (MCA) degree in 2004 from P.S.G.R.Krishnnammal College for Women, Coimbatore, India. She received M.Phil (Computer Science) in 2006 from Bharathiyar University. She received Ph.D.(Computer Science) in January 2023 from Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore. Her research interest is Data Mining andMachine Learning. She haspublished 8 papers in International Journals out ofwhich four are indexed journals and 4 other papers in International Conferences as apart of her research work.

Kirsten Kennedy is a MCA Student,studying II year in the Department of Computer Applications, Kumaraguru College of Technology, Coimbatore, India.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details