Rainfall Prediction Using Machine Learning

Santosh Singh[1], Vishal Pandey[2], Abhishek Singh[3]

[1]Head of Department (Information Technology), [2,3]P.G. Student (Data Science)

[1,2,3]Thakur College of Science & Commerce, Thakur Village, Kandivali East, Mumbai 400101, Maharashtra, India

Abstract

Agrarian nations like India are highly dependent on rainfall amounts to determine the success or failure of agriculture. Most of India's agricultural produce relies on the monsoon rains and its precipitation patterns. Monsoon season provides for most water needs in India. For adequate crop planning, it is important to know average rainfall. Several experiments say that precipitation directly affects crops. Therefore, this study used machine learning techniques and algorithms to analyse Indian states' rainfall patterns using rainfall data from government during 1901-2017 period. The study concludes that machine learning algorithms are useful in analyzing and investigating about Indian states' rain fall with comparable results to those in industry standards.

Keyword - Rainfall Prediction, Machine Learning Model, Regression Analysis, Random Forest, Data Preprocessing, Accuracy Assessment

Introduction

Determining the amount of rain and its forecast is highly essential in weather prediction, modern living, agriculture, and other ecosystems. Farm planning and effective water resources management are based on understanding the rainfall as well as predicting it. The problem here lies that the output must be perfect. For instance, there are various hardware devices to predict rain using weather conditions like; pressure and temperature. It is inefficient through these approaches; therefore, we need machine learning techniques to give exact results. By just doing analysis on the past records of precipitation one can establish this year's rainy season. Some methods have different accuracies hence one should select an algorithm with a view to meeting its model requirements. Better machine learning methods for higher rainfall forecasts accuracy. The present study uses Indian Sub-Divisional Meteorological Data Set from 1901 to 2017. There were multiple subdivisions for data collection in meteorology. however examining with high precision and forecasting rainfall trend. Some software tools including xgBoost, linear regression and multiple linear regression were applied in our experiment.

Literature Review

A study on rainfall pattern in Indian states using machine learning techniques and algorithm was done by Nikhil Tiwari et.al. The dataset comprised of sub-divisionally monthly precipitation and its departure from 1901 to 2017. It contains precipitation value of the month when it rains most in India such as October, November etc. This model aims at predicting month-wise and year-wise precipitation (In mm). For all these Linear Regression, Random Forest Regression & Elastic Net So far as accuracy is concerned, amongst them Linear Regression can achieve 94.0% highest precision than any other algorithm. Temporal variability in the

pattern of rain is significant. However this complexity might not be understood well enough to reflect variations in climates[1].

B.Suganya et.al suggested a rainfall prediction model using Multiple Linear Regression for Indian dataset.The input dataset contains several meteorological parameters& would predict the rainfall more accurately. Precipitation Estimation Quantitative Precipitation Estimate, Linear Regressions & Multiple Linear Regressions In all these algorithms Multiple linear regression could show a better performance with Correlation 0.473 being higher than other methods "dataset don't cover a widge extend of climate condition over diverse region in India[2].

Moulana Mohamad et.al conduct a research on prediction of rainfall using machine learning techniques. The dataset consist of sub divisionally monthly rainfall & its departure from 1901-2015. The attributes are the amount of rainfall measured in mm. In this use Multiple Linear Regression, Support Vector Regression & Lasso Regression In all of these Support Vector Regression can achieved 99.2% most accuracy then other algorithm. The accuracy of prediction Literally relies on the quantity & quality of historical weather data. If you do not have accurate data, the prediction may not be dependable[3].

Yousisf Ahmed et.al proposed rainfall prediction using multiple linear regression model. The rate of precipitation, or rainfall rate, for Khartoum state has been predicted by the authors of this research using Multiple Linear Regression model, based on weather terms that are considered the independent variables. Assumptions of Linear Regression: Multiple Linear Regression models a linear relation between the dependent variable (rainfall) & the dependent variables(predictors).Although, rainfall pattern can exhibit nonlinear behaviour's, mainly in complex climatic System. This can lead the a inaccurate forecast[4].

Sowmya V et.al conducted research using rainfall prediction using machine learning & deep learning techniques. Rainfall is essential for India agriculture, preventing flooding. Although predicting rainfall is challenging due to fluctuations. Researcher use various techniques like fuzzy logic, Artificial Neural Network, regression to create accurate models. Deep learning models that require a lot of resources for training, like CNNs & RNNs can be computationally taxing. Because of this, it may be challenging to scale them for large dataset with limited resources[5].

Methodology

The predictive model is used to prediction of the precipitation. The first step is converting data in to the correct format to conduct experiments then make a good analysis of data and observe variation in the patterns of rainfall. We predict the rainfall by separating the dataset into training set and testing set then we apply different machine learning approaches (XGBoost, MLR, RF, Lasso etc.) and statistical techniques and compare and draw analysis over various approaches used.

Dataset Description:

You have collected dataset from Indian Metrological Department from 1901-2017 from Sub division of Indian states. Data set consist of 15 attributes(individual months, annual) for 36 Sub division of country in which pattern of rainfall are observed and collected. The attributes are measured pattern of rainfall in mm. As the dataset is very large, feature reduction is done so that it Improves the accuracy, reduces the computation time and also storage. Exploratory Data Analysis (EDA) is techniques to learn about the distributional pattern. prior to EDA data cleaning process was executed to handle missing values, outliers and Utilize correlation Analysis, recursive feature elimination. Descriptive statistics and visualization use to identify anomalies and guide the cleaning process.

XGBoost : XGBoost is a powerful machine learning algorithm that is commonly used for classification and regression task. According our dataset it deal with a regression problem where goal is to predict a continuous numeric target variable. It work on parallel computing and distribute the work load on multiple cores.

Multiple Linear Regression : Multiple Linear Regression is a statistical method used to analyze the relationship between two or more independent variable (features) and a dependent variable (rainfall). The mathematical expression of multiple linear regression is

$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n + \varepsilon$

where Y represent dependent variable, b_0 is the intercept and $b_1, b_2, \ldots, b_n$ are coefficient and $X_1, X_2, \ldots, X_n$ represent independent variable , ε is a error term. This algorithm is try to find minimizing the sum of square differences between the predicted and actual values.

Random Forest Regression: Random forest is based on ensemble learning techniques. Random Forest produce result by combining several Decision Trees results which run in parallel without affecting one another. It is versatile algorithm that often perform well in various prediction tasks including stock market and rainfall prediction. It can handles high dimensional data well, provides estimates of feature importance.

Lasso Regression : Lasso Regression is a linear regression techniques. This algorithm performs two methods one is variable selection and other is regularization to improve the predication accuracy. The Lasso method incorporates a penalty term that shrink certain coefficient to zero.
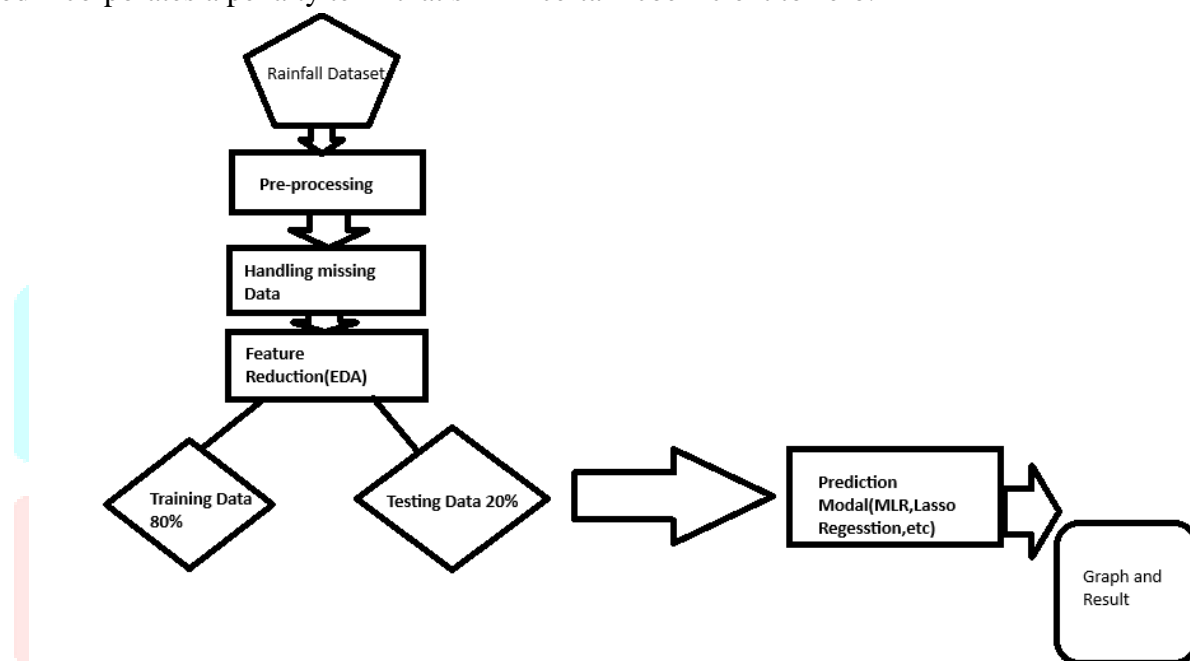

Fig 1. Rainfall Prediction Model.

Result

Four distinct machine learning algorithm can be evaluated: Random Forest, XGBoost, Lasso Regression and Multiple Linear Regression. As far as the correlation between the variables is concerned, fig.2 sows the high correlation between some variables, such as June to July. Each modality assessed was evaluated based on the R score, which indicates the proportion of variance in an independent variable that is predictable from the dependent variable. Although, there was a really low point from January to July. February through October are the months with the opposite rainfall patterns.
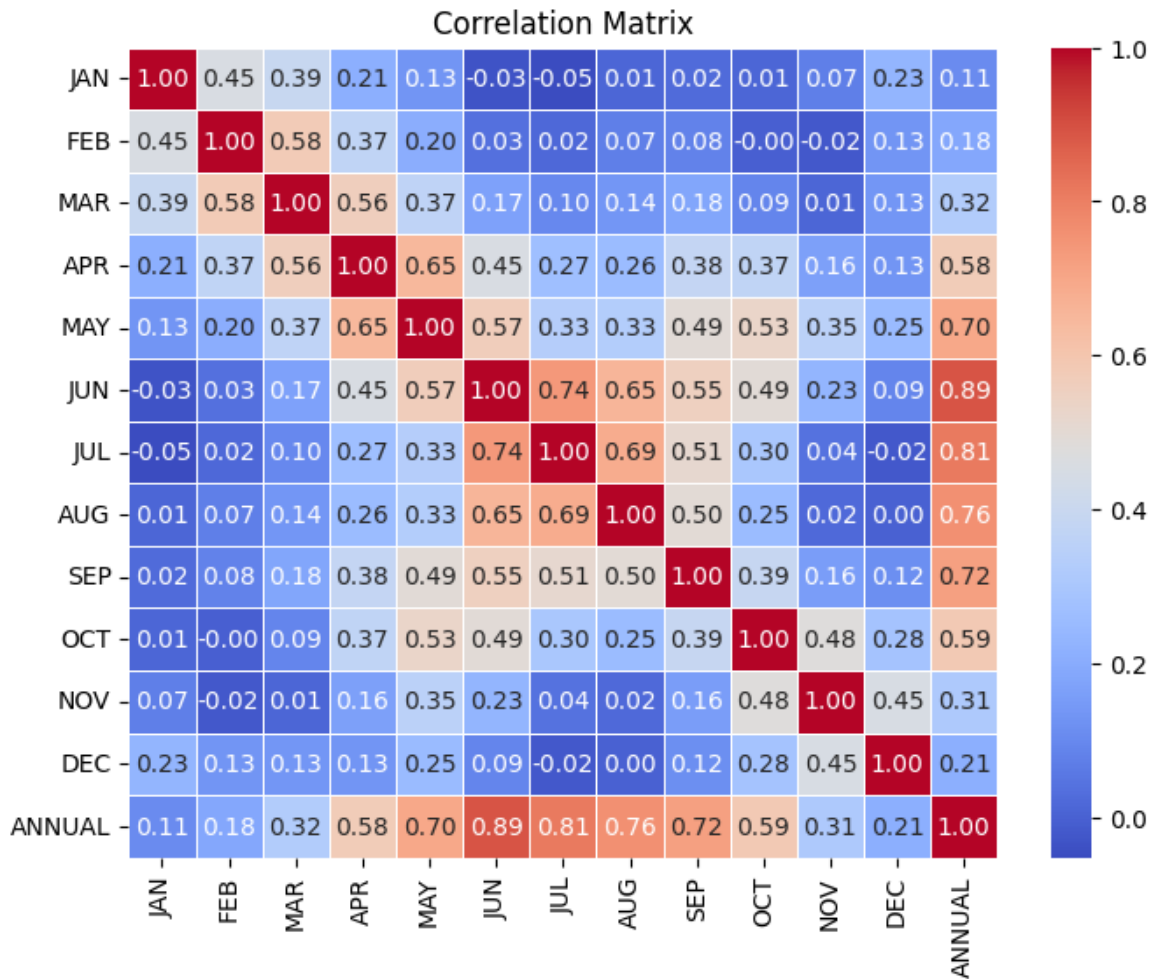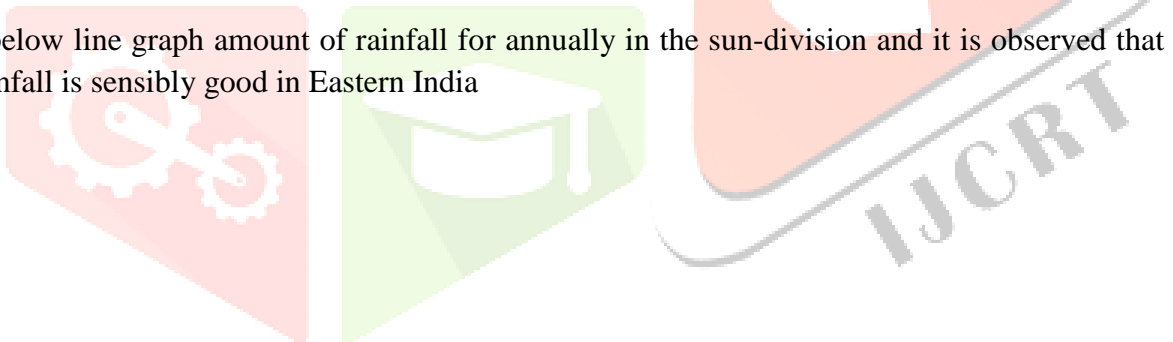
Fig 2. Monthly Correlation

The below line graph amount of rainfall for annually in the sun-division and it is observed that the volume of rainfall is sensibly good in Eastern India
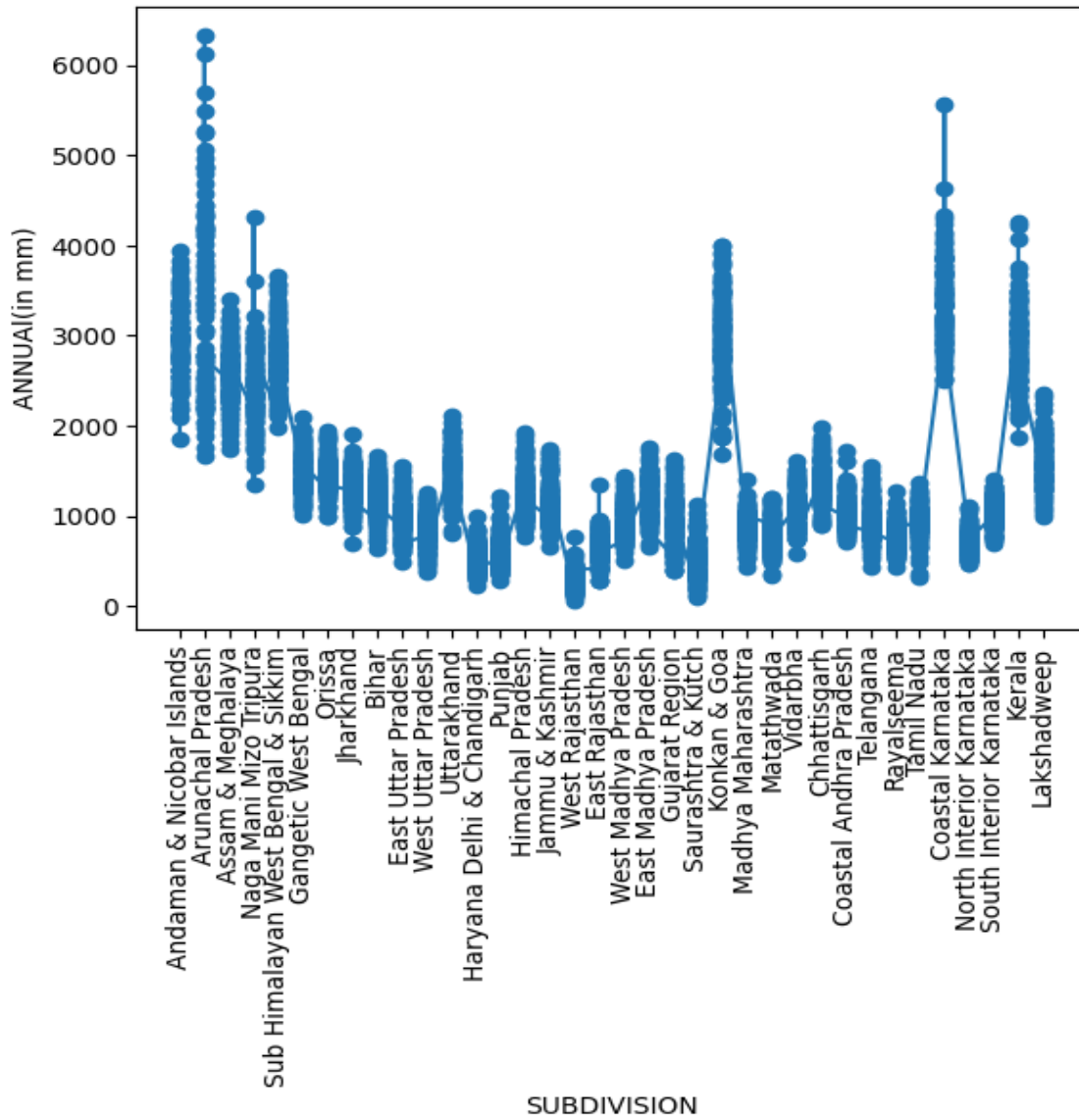
Fig 3 Line graphs for the amount of rainfall in all subdivision annually.

Modal	R^2
XGBoost	0.9819
Lasso Regression	0.9910
MLR	0.9950
Random forest	0.9737

Table 1: Accuracy of the model

Conclusion

This research purposed a study a rainfall in India and a productive ananlysis of various state of machine learning algorithm which include XGBoost, Multiple Linear Regression, Random Forest Regression, Lasso Regression and evaluated them most suitable metric i.e. R^2 score. This attempt was made to improve the previous result by using ensembled machine learning algorithm.. Overall, this study's Advanced Rainfall Prediction Accuracy highlights the need for knowledge when making decisions about agriculture and water resources.Furthermore robust, we have ensemble techniques, methodology, and other machine learning algorithms to enhance our forecast powers in tackling the intricate dynamics of our rainfall pattern

References:

1] N. Tiwari and A. Singh, "A Novel Study of Rainfall in the Indian States and Predictive Analysis using Machine Learning Algorithms," *2020 International Conference on Computational Performance Evaluation (ComPE)*, Shillong, India, 2020, pp. 199-204, doi: 10.1109/ComPE49325.2020.9200091.

2] R. K. Grace and B. Suganya, "Machine Learning based Rainfall Prediction," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 227-2210.1109/ICACCS48705.2020.9074233.

3] Mohammed M, Kolapalli R, Golla N, Maturi SS. Prediction of rainfall using machine learning techniques. International Journal of Scientific and Technology Research. 2020 Jan;9(01):3236-40.

4] H. A. Y. Ahmed and S. W. A. Mohamed, "Rainfall Prediction using Multiple Linear Regressions Model," 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan, 2021, pp. 1-5, doi: 10.1109/ICCCEEE49695.2021.9429650

5] C. Z. Basha, N. Bhavana, P. Bhavya and S. V, "Rainfall Prediction using Machine Learning & Deep Learning Techniques," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 92-97, doi: 10.1109/ICESC48915.2020.9155896