International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

USING MULTIPLE LINEAR REGRESSION MODEL FOR RAINFALL PREDICTION IN INDIA

[1]C.VIDYA, [2]SD.ASLAM BASHA, [3]B.SAI KIRAN, [4]SK.YOUNUS BASHA, [5]U.BALA VENKATESH, [6]SD.MUZAKKIR

[1]Assistant Professor, Department of ECE, Sree Venkateswara College of Engineering, North Rajupalem(VI), Kodavaluru(M), S.P.S.R Nellore (DT), Andhra Pradesh, India.

[2,3,4,5,6]B.Tech Scholars, Department of ECE, Sree Venkateswara College of Engineering, North Rajupalem(VI), Kodavaluru(M), S.P.S.R Nellore (DT), Andhra Pradesh, India.

ABSTRACT: **India is as an agricultural country where crop productivity contributes to a major share in the economy. For understanding the crop productivity, prediction of rainfall is required and necessary. Forecasting is a challenging work. Weather condition is the state of atmosphere at a given time in terms of weather variables like rainfall, cloud conditions, temperature, etc. The occurrence of rainfall is an outcome of various natural factors such as temperature, humidity, cloudiness, wind speed, etc. Rainfall prediction is a major concern for meteorological department as it is closely associated with the economy and sustenance of human life. This analysis describes to develop a Multiple Linear Regression (MLR) model in order to predict the rate of precipitation (PRCP), i.e., rainfall rate. It is based on some weather parameters, such as temperature, wind speed, and dew point. The efficiency of the model has been measured by comparing the average value of the mean square error of the training data with the test data. The obtained results show that the average of the mean square error has been improved.**

KEYWORDS: Weather Prediction, rainfall, Linear Regression, Machine Learning, Rainfall Prediction

I. INTRODUCTION

In Today's era global warming is affecting all over the world which majorly effect on mankind and cause the expedite the change in climate. Due to this air and oceans are warming, sea level is rising and flooding and drought etc. One of the serious consequences due to this climate change is on Rainfall. Rainfall prediction now days is an arduous task which is taking into the consideration of most of the major world-wide authorities. Rainfall is a climatic factor that affects several human activities power generation and tourism, among others. This make the rainfall serious concern and requirement of better rainfall prediction. Rainfall is a complex atmospheric process, and due to the climate changes, it becomes more difficult to predict it. Since the arbitrary characteristics of rainfall series, they are often labelled by a stochastic process.

Rainfall is a natural phenomenon defined as the outcome of interaction between several complex atmospheric processes. A large uncertainty involved in determining the contribution of the atmospheric processes is one of the biggest challenges to face in developing rainfall prediction models. Rainfall prediction is very difficult to model since the atmospheric processes involved follow a rather complex nonlinear pattern. The temperature, relative humidity, wind speed, wind direction, cloud coverage etc. are some of the factors that critically affect the occurrence of rainfall.

The process of predicting the state of the atmosphere for a specific location in the future is called weather forecasting [3]. Interest in weather forecasting began from the earliest era, and the forecasting techniques were developed and have been changing with time, Several methods are used to generate weather forecasting, each of which differ in its accuracy and efficiency. There are three important steps that must precede the process of weather forecasting, which are to collect

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

atmospheric data as much as possible, to understand the data and its inter-relation to determine the behavior of the atmosphere, and to use it in numerical models to predict the future state of the atmosphere. Recently, scientists tended to apply machine learning tools for weather prediction, because it does not require a deep and comprehensive understanding of the atmospheric process, thus it represents a good choice for weather forecasting [5].

Machine learning (ML) is a process of learning a specific task without any human intervention, which will improve the performance only by the continuous learning process. Learning methods are of three types: supervised learning that is based on labeled data, unsupervised learning, and the reinforcement. The vital process in all machine learning methods is extracting of the features, and then to use these.

The extracted features for various approaches, like classification and regression [1]. Applying machine learning techniques in weather forecasting can compensate complex meteorological physics model. With the availability of metrological data set, the two authors were encouraged to select supervised learning method, which is multiple linear regression, instead of unsupervised learning or reinforcement learning. There is different regression types used in machine learning, such as linear regression, logistic, polynomial regression. The simpler and most frequent method is linear regression, which is used for prediction [7].

II. LITERATURE SURVEY

F. Olaiya and A. B. Adeyemo Olaiya, et.al [11] used an Artificial Neural Network and Decision Tree algorithms and meteorological data (2000- 2009) for the city of Ibadan, Nigeria, in forecasting weather variable (maximum temperature, rainfall, and wind speed). S. Prabakara, P. N. Kumar, and P. S. M. Tarun, et.al [4] used a modified linear regression model to predict rainfall with less error percentage by adding percentage to the input values.

S. M. Paras, et.al [6] used the Multiple Linear Regression (MLR) model to predict four weather parameters which are (maximum and minimum temperature, relative humidity, and the category of rainfall). W. M. Ridwan, M. Sapitang, A. Aziz, K. F. Kushiar, A. N. Ahmed, and A. El-Shafie, et al [2] have applied two methods to predict rainfall forecasting rainfall, which are Autocorrelation Function (ACF) and projected error. Both methods implemented four different regression algorithms (Bayesian Linear Regression, Boosted Decision Tree Regression, Decision Forest Regression and Neural Network Regression, with different time horizons (daily, weekly, ten days and monthly),The results showed that Boosted Decision Tree Regression is the best regression developed for M1, with the highest coefficient of determination, but in M2 the overall model performance gives a good result of each category except for 10-days with Boosted Decision Tree Regression and Decision Forest Regression.

Daniela Şchiopu, Elia Georgiana Petre, Catalina Negoina, et.al [14] 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 R. Samuel Selvaraj and Raajalakshmi, et.al [15] used multiple linear regression to predict the monsoon rainfall by using outgoing long wave radiations, global temperatures and sunspots out of Tamil

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Nadu. They collected data from 110 years from Indian Meteorological department, Chennai.

Dhawal Hirani and Dr. Nitin Mishra, et.al [8] 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 Mathur, et.al [12] 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.

M. Kannan, S. Prabhakaran, P. Ramchandran, et.al [13] implemented Multiple linear regression and Karl Pearson coefficient. They made a short – term forecast over aparticular state. They used fuzzy sets, neural networks to analyse the data.

Delson Chikobvu and Retius Chifurira, et.al [10] 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 also proposed the F-test and Screening procedure. a cross-validation procedure is used first to screens models out that are all likely to Poorly perform 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, et.al [16] 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 Sethi et al, [9] 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.

III. USING MULTIPLE LINEAR REGRESSION MODEL FOR RAINFALL PREDICTION IN INDIA

In this section, using multiple linear regression models for rainfall prediction in India is discussed. The data of rainfall and precipitation of all the states of India for a period of 64 years from 1951-2014 was taken from government meteorological site. Multiple linear regression tries to show the connection between at least two illustrative variables and a response variable by fitting a linear equation to experiential data.

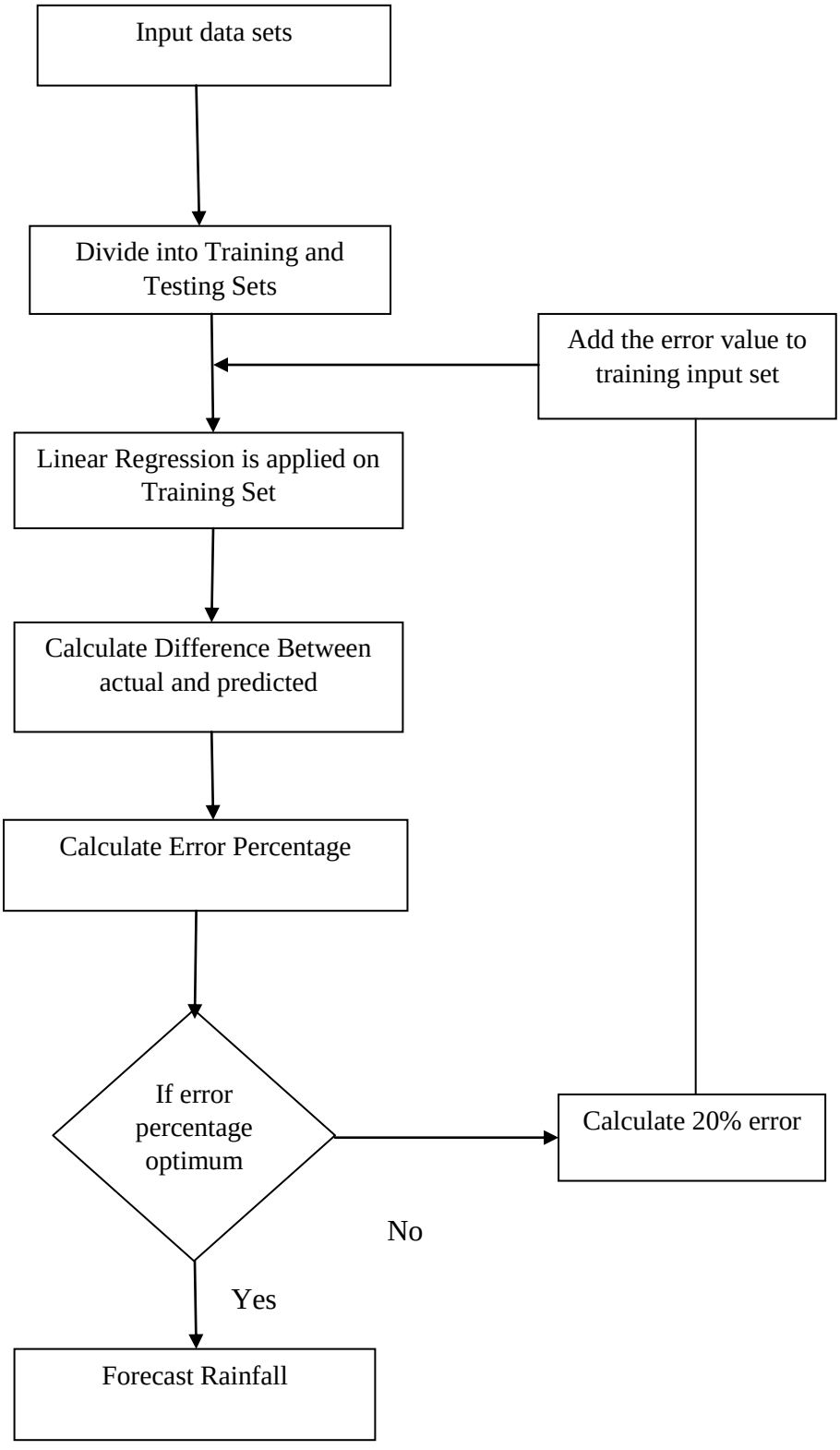
The input to this model are rainfall and precipitation and variables that are extracted from the dataset using correlation. The MLR model will be used as the Predictive models to predict rainfall. The structured rainfall and precipitation data in the form of csv file or excel file will be taken as the input for the model. R language is used to create the predictive model.

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. 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

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

data from 1971 to 2002 on which the testing of model is done.

Fig.1: Block Diagram



International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

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.

Linear regression is one type of the supervised learning techniques to predict a numeric value (dependent variable) form a set of features (predictors). Likewise, it is about finding a function that maps inputs x € R to the corresponding function values f (x) € R, It forms a prediction by computing a weighted sum of the input features, plus a constant called the bias (intercept), When the dependent variable is calculated from one predictor, the regression is called simple regression, as shown in Equation (1) below.

$$Y = a + bX \quad (1)$$

Where,

Y: dependent variable a: intercept
b: slope
X: independent variable

If it is produced from two or more predictors, the regression is called multiple regressions, as shown in Equation (2) below.

$$Y = a + b1\ x1 + b2\ x2 + \ldots bn\ xn \quad (2)$$

Where,
x1 , x2 , ... xn : independent variables.

As other supervised learning methods to develop the linear model, authors have passed through two phases: the training phase also named the learning phase [6] and the testing phase. At the training time, they used a well- defined labeled data to adjust the bias and features weight to obtain multiple linear regression equations. At the testing phase, they used another labeled data so as to verify the validity of their model as a generalized model for prediction.
the predicted value from the actual value and multiplying it with 100 to get percentage. 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. 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. 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. 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. In the above steps the coefficients of the independent variables in the linear regression keep on changing. The latest updated coefficients to forecast the test data and this produces the most accurate forecast values.

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.

IV. PERFORMANCE ANALYSIS

In this section, performance analysis of multiple linear regression model for

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

rainfall prediction in india is discussed here.

Table.1: Performance Analysis

Parameters	MLR	ARIMA
Accuracy	98	91
Mean Square Error	0.3	0.18

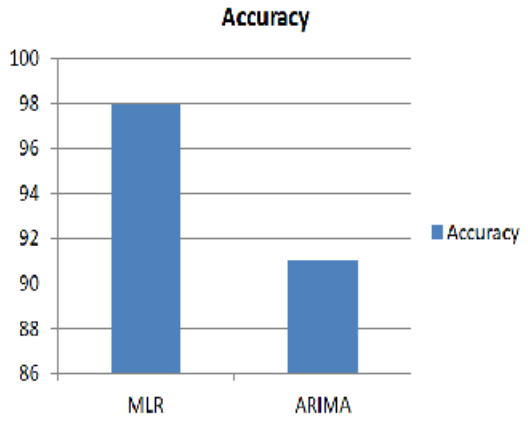


Fig.2: Accuracy Comparision Graph

In fig.2 accuracy comparision graph is observed between MLR and ARIMA. MLR shows higher accuracy.

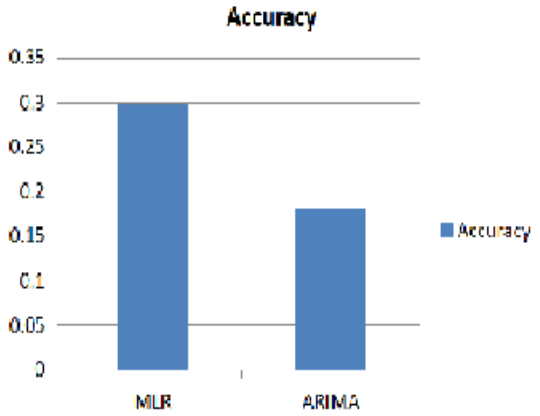


Fig.3: Mean Square Error Comparision Graph

Mean Square Error of MLR shows higher when compared with ARIMA in fig.3.

V. CONCLUSION

Rainfall is the major cause for many of the natural disasters like flash floods, droughts, tsunamis. So in order to prevent these natural calamities. The 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 system uses multiple linear regression approach to predict the rainfall that has less error percentage than compared to ARIMA. This data is used to perform the necessary calculations to predict the rainfall from average temperature and cloud cover of that particular district.

VI. REFERENCES

[1] J. Refonaa, M. Lakshmi, R. Abbas, and M. Raziullha, "Rainfall Prediction using Regression Model," International Journal of Recent Technology and Engineering *(IJRTE)*, vol. 8, no. 2S3, Jul. 2019.

[2] W. M. Ridwan, M. Sapitang, A. Aziz, K. F. Kushiar, A. N. Ahmed, and A. El-Shafie, "Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia," Ain Shams Engineering Journal, 2019

[3] E. Abrahamsen, O. M. Brastein, and B. Lie, "Machine Learning in Python for Weather Forecast based on Freely Available Weather Data," Proceedings of The 59th Conference on imulation and Modelling (SIMS 59), 26-28 September 2018, Oslo Metropolitan University, Norway, 2018.

[4] S. Prabakara, P. N. Kumar, and P. S. M. Tarun, "RAINFALL PREDICTION USING MODIFIED LINEAR REGRESSION," ARPN Journal of Engineering and Applied Sciences, vol. 12, no. 12, Jun. 2017.

[5] M. Holmstrom, D. Liu, and C. Vo, "Machine Learning Applied to Weather Forecasting," Dec. 2016.
[6] S. M. Paras, "A Simple Weather Forecasting Model Using Mathematical Regression," Indian Research Journal o f Extension Education, vol. 12, pp. 161-168, 2016.
[7] S. Gupta, I. K, and G. Singhal, "Weather Prediction Using Normal Equation Method and Linear regression Techniques," International Journal of Computer Science and Information Technologies, vol. 7, no. 3, pp. 1490-1493, 2016.
[8] 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.
[9] Nikhil Seth, " 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
[10] 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. Vol. 5 No. 7 (2014): May 2014
[11] F. Olaiya and A. B. Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies," International Journal o f Information Engineering and Electronic Business, vol. 4, no. 1, pp. 51-59, 2012.
[12] 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, Volume 6– No.2, March- April 2012.
[13] M. Kannan, S. Prabhakaran, P. Ramachandran "Rainfall Forecasting Using Data Mining Techniques". International Journal of Engineering and Technology Vol.2 (6), 2010, 397-401.
[14] 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.
[15] 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, 2009.
[16] Guhathakurta, (2005) "Long-range monsoon rainfall prediction of 2005 for the districts and subdivision Kerala with artificial neural network", Current Science, 90, 773-779.