Using Machine Learning Approaches to precisely Predict RainFall

Chinta Gouri Sainath P [1], Anvesh Kumar A [2], Naveen Kumar A [3]

[1] Assistant Professor, Department of Information Technology
[2] Assistant Professor, Department of Computer Science and Engineering
[3]Assistant Professor, Department of Computer Science and Engineering
[1] CMR College of Engineering and Technology,Medchal,Hyderabad,India.
[2] Malla Reddy College of Engineering and Technology,Kompally,Hyderabad,India.
[3] Malla Reddy College of Engineering and Technology,Kompally,Hyderabad,India.

Abstract- Meteorologists continually seek techniques to grasp the Earth's atmosphere and build precise weather forecast models. In weather prediction, several approaches were employed. Machine approaches are recently thought to be precise procedures and have been widely utilised for weather prediction, as an alternative to conventional methods. The precipitation rate has an important impact on agriculture and the bio-sector in the weather system. This article aims at the development of multiple linear regression model to forecast the rainfalls (PRCP). It depends on several meteorological characteristics such as temperature, wind velocity and wind speed. The information utilised in this research is published on the National Climate Data Center website. A Python code was created in order to develop the pattern for artificial neural networks using the package Pytorch. Measured by comparing the mean square error value of the training data to test data, the efficiency of the model has been measurable. The findings achieved demonstrate that, while using the same amount of data throughout trainings and testing periods, the average square error has been improved by 85 percent during trial time. However, the amount of data during the test phase goes above the quantity of data in the training phase to 59 percent.

Keywords—Brain Training, Wind speed, Linear regression, Weather forecasting, Training data, Terrestrial atmosphere, Mean square error methods

I. INTRODUCTION

The prediction of rebound is one of the most significant methods for predicting the weather of all countries. In the Indian civilization, the prediction of rebounds has a vital function to play in human existence. The meteorological department is responsible for predicting the frequency of precipitation and for uncertainties. With shifting climatic circumstances, predicting precipitation properly is hard. The precipitation during both the summer and the rainy season is difficult to predict. Researchers across the world have created several models to forecast precipitation using largely random values [1]. Multiple linear regression is used to build the suggested model. The approach suggested uses Indian weather data to forecast the fall of rain. Machine learning algorithms are usually divided into two main categories: (*i*) unsupervised learning (*ii*) supervised learning. All the group algorithms are monitored by the computer. Although numerous models have evolved [2], it is important for study to accurately forecast utilizing machine-learning algorithms. Error-free prediction gives farmers and other businesses better planning. The following document is organized: The many relevant literature techniques are discussed in Section II. The suggested MLR based Rain Fall Prediction is described in Section III. In Section IV, the results are produced and in Section V the paper ends.

II. RELATEDWORKS

S. K. Sunori et al., (2021) [3]. Their research is the precipitation with knowledge of two rainfall factors, i.e. temperature and humidity which are significantly linked to the precipitation. The predicted rainfall in India is a difficult assignment since the weather here is substantially changing. Their study utilizes techniques of artificial intelligence (AI) to develop a prediction model for the prediction of precipitation amounts. Temperature and humidity are the two considered input factors for this model. Two distinct AI techniques have been developed, the first one is subtractive and the other Levenberg-Marquardt algorithm. The MATLAB predictive model is used. Finally, their performance forecast is taken into account.

N. Parashar et al., (2021) [4] Their study aims to offer a thorough assessment of recent scientific work on local and global spatial predictions of short-term temperatures and precipitation. Some weather conditions such as air pressure, precipitation, wind speed, wind direction and temperature at the dew point, solar radiation, vapour pressure, cloud cover, skies or snow are possible measures for predicting the future temperature. The literature shows that the weather conditions are not unconscionable. It focuses on contemporary apps for machine learning and deep learning models such as the Deep Echo State Network, recurrent neural networks, recurring neural networks and graph-based co-operation networks, auto encoders, perceptrons and long-term memory applications. Multimodal study applications, reservoir computing and multi-task learning have demonstrated significant improvements in the prediction accuracy of other state-of-the-art models. CNN is also considered to be fine at extracting appropriate patterns from numerical weather data. The data recording time period also significantly influences the accuracy of the forecast. The input data recorded more often functioned better than the data recorded less often. Instead of satellite and radar installations the use of electromagnetic sensors is dependable and economical for data collection to anticipate. Assessment indices associated with the rainfall impact and no rainfall, caching rate, the overview rate and the swing-and-miss rate can, together with other statistical metrics in the rainfall preview, be treated as statistical measures.

U. Ashwini, et al., (2021). [5] The number of predictions of rainfall is an important problem for the meteorological service because of the human life and economy. Excessive precipitation is the main cause of naturally occurring calamities like drought and floods that people worldwide face every year. For predicting Tamilnadu precipitations, the time series machine learning model is utilised. The Indian meteorological agency can provide forecast data needed for analysis. The seasonal ARIMA (Auto Regressive Integrated Movement Average) approach is utilised to model monthly precipitation in Tamilnadu from January 1990 to December 2017. The stationarity of time series flow was proven by the prediction model for the rainfall and the evaluation of the seasonal correlogram using the SARIMA (Seasonal Auto Regressive Integrated Moving Average). The output of this model is

evaluated in reference to the mean square error (MSE) and routine mean square error (RMSE). The ARIMA model therefore indicates that Rainfall with fewer error is successfully forecasted and the model developed might be utilised to predict Monsoon rainfalls in the next years.

N. Barclay et al., (2021). [6] In recent years, extreme rainfall has been increasing and intensifying, resulting to rising floods. Possible consequences are amplified due to their age, particularly on transport infrastructure such as highways and bridges. Transport infrastructure is also struggling to keep up of rising land expansion, leading to greater stream flow. A software tool is suggested that contributes to decision-making in disaster response for heavy rainfall events to minimise the impact of bridge overtoping and road washout on the public. The specified research region is developed utilising QL2 LiDAR data sets for creating the software tool by means of a precise, high resolution surface topography model (containing canals, roads and bridge and culvert infrastructure). Furthermore, model storm-water flows for flow at flux and crossing points and through culverts during different storm return times are created. The resultant software tool will incorporate location identification for bridges overtopping and road washout to demonstrate the sevenfold impact in almost real time. This tool is available as open source (for example, Python) software and an interactive map that shows the locations and severity levels based on planned rain data input for the field of research. This paper first introduces the approach presented, describes the processing and collecting of data, and then analyses a case study. The findings confirm the first stages in data gathering and processing to promote future analysis and modelling.

S. Soni, et al., (2021). [7] To anticipate the future weather conditions, the probability will be lower than the previous day's prediction, but the likelihood will be high in the following two week's. The two weeks weather data from the slide window of the previous year must thus be processed to select a size equivalent to one week. The present year overlaps with every rapid window week. Moreover, a window algorithm slide is used to forecast. The findings of the technique show that, with an average accuracy of 94,2 percent, the use of the suggested weather prevention strategy is successful. While the radar remote sensing sector for PWS is one of the future technical developments that will be most interesting and innovative. In addition, the next-generation radar systems (dual-polarization radar, phased-array radar) would improve extreme weather, rainfall predictions and winter weather alerts while enhancing lead times for severe threats such as tornadoes, and heavy rain/flash floods.

III. PROPOSED SYSTEM

Universal prediction procedures comprise data collection, pre-processing and selection of data, reduction of explanatory predictor, regressive constructing model and last validity control as illustrated in Figure 1. The first crucial stage in data mining is data collection. The Weather data collection is gathered by the Department of Metrology in India. The department keeps the data sheet in both monthly and annual form. In data mining, data preparation is the next hard

task, the data gathered till now is loud and some missing values and undesirable data are present. The data must be cleaned and irrelevant data removed by the collection of missing values. After data pre-processing, the selection of data is the next stage in this regard, we must pick the data that is important to our analysis and leave all other information that we use to decide which correlate. Then predictors with a strong interrelationship with others are removed since the existence of several highly interrelated explanatory factors can significantly enhance samples and impair the predictive model capacity of the coefficients. The next stage is the constructing model using training data following the explanatory predictors for reduction. The method employed here is a technique of linear regression.

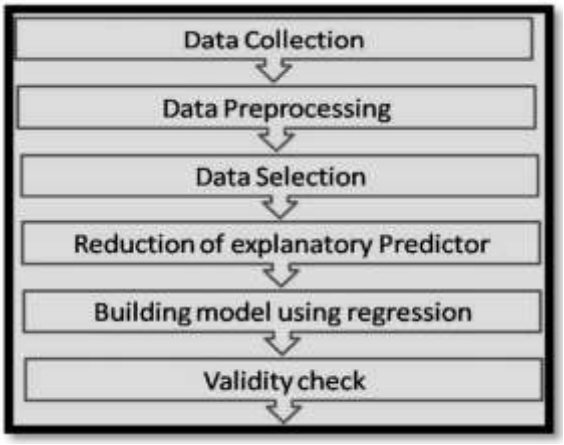

Figure 1. Model for test data verification.

IV. RESULTS AND DISCUSSION

The tests were conducted to assess the precision of precipitation with multiple linear regressions. In this section, the forecast results are shown. The projected rainfall level is compared to the actual rainfall to assess the quality of the MLR equation. Regional precipitation data collected from India and precipitates are utilized as predictors, for experiments like cloud cover, average temperature and vapor pressure. For the experiment the data set is utilized for 30 years. Details of the predictor association with predictive precipitation are shown in the table I.

Table I. Correlation of predictor with rainfall

SL. No.	Predictor	Correlation Coefficient
1	Precipitation	0.315
2	Cloud cover	0.582
3	Average temperature	-0.285
4	Vapor pressure	-0.260

There are several methods, such as linear regression, log based regression and nonlinear prediction regression. We utilized many regression approaches for the data set here. In any future year, we can predict the precipitation using climate parameters from this technique. We now choose four climatic variables with rain datasets from India to proceed with this technique. Applying different techniques to regression on this data set and identifying predictable rain- and climate equations. Therefore the MLR below is specified.

Y=-1323.062+0.237*X10.75*X2+16.317*X3+13.038*X4

Where

Y=Predicted rainfall

X1=Precipitation

X2=Average Temperature

X3=Cloud Cover

X4=vapor pressure

From that equation, the precipitation, the mean temperature, cloud cover and the vapour pressure need to be calculated for future years. The figure 2 shows the graphic depiction between the actual and the anticipated value of the rainfall when the MLR equation is applied with test data to test the accuracy of the MLR equation.

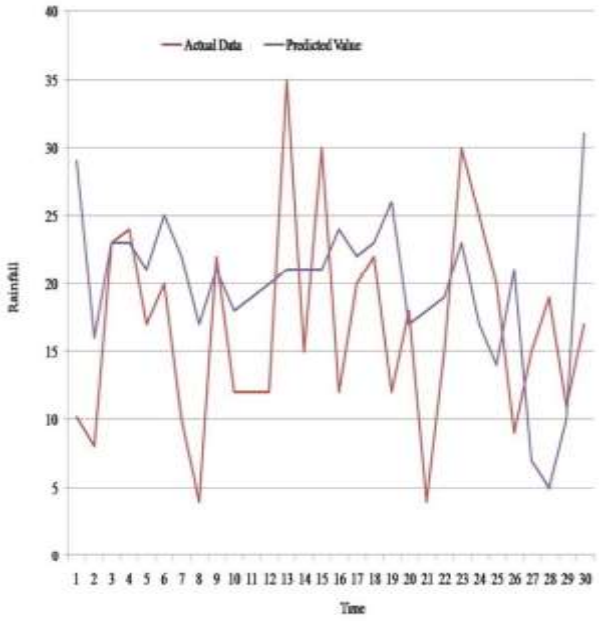

Figure 2. Comparison among actual data and predicted value.

V. FUTURE SCOPE AND CONCLUSION

After analyzing Indian precipitation data set, we picked a method of precipitation prediction that was developed from certain ML methods, for example, correlation analysis and regression analysis. Rainfall has a major influence not just in India but throughout the world on agriculture, on the economy. In order to anticipate rain in the next year, we are familiar with the climate parameters that are quite important for agriculture. This is the sole rain forecast, although not exact due to climatic conditions. As we know, we have been using some elements to affect the rain, and we know the factors that fluctuate according to different reasons.

REFERENCES

[1] 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), 2021, pp. 1-5, doi: 10.1109/ICCCEEE49695.2021.9429650.

[2] F. D. Diba, M. A. Samad and D. -Y. Choi, "The Effects of Rain on Terrestrial Links at K, Ka and E-Bands in South Korea: Based on Supervised Learning," in IEEE Access, vol. 9, pp. 9345-9355, 2021, doi: 10.1109/ACCESS.2021.3049825.

[3] S. K. Sunori et al., "Rainfall Prediction using Subtractive Clustering and Levenberg-Marquardt Algorithms," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 1458-1463, doi: 10.1109/ICOEI51242.2021.9452869.

[4] N. Parashar and P. Johri, "Short-Term Temperature and Rainfall Prediction at Local and Global Spatial Scale: A Review," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 742-746, doi: 10.1109/ICACITE51222.2021.9404767.

[5] U. Ashwini, K. Kalaivani, K. Ulagapriya and A. Saritha, "Time Series Analysis based Tamilnadu Monsoon Rainfall Prediction using Seasonal ARIMA," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1293-1297, doi: 10.1109/ICICT50816.2021.9358615.

[6] N. Barclay and M. Smith, "Assessment of Advanced LiDAR Based Tools for Enhanced Flood Prediction," SoutheastCon 2021, 2021, pp. 1-5, doi: 10.1109/SoutheastCon45413.2021.9401895.

[7] S. Soni, K. Vashishtha and C. Bhandubey, "Deep Learning based Weather Forecast: A Prediction," 2021 5th International Conference on Trends in

Electronics and Informatics (ICOEI), 2021, pp. 1736-1742, doi: 10.1109/ICOEI51242.2021.9452855.