# Effective Approach for Predicting Rainfall Using Machine Learning

Mr. M. Ravindra Babu[1], V. Sai Harini[2], M. S. Subrahmanyam[3], Reena Reddy. K[4], L. Padma[5]

[1]*Asst. Prof, Dept of CSE, Tirumala Engineering College, Narasaraopet, Guntur, A.P., India*

[2,3,4,5]*B. Tech Students, Dept of CSE, Tirumala Engineering College, Narasaraopet, Guntur, A.P., India*

**ABSTRACT -** Rainfall forecasting is critical because heavy rains can cause a variety of disasters. The forecast enables people to take proactive steps, and it should also be reliable. There are two forms of rainfall prediction: short-term rainfall prediction and long-term rainfall prediction. Predictions, particularly short-term predictions, can provide us with accurate results. The key challenge is to create a model for predicting long-term rainfall. Heavy precipitation prediction may be a big disadvantage for the earth science department because it is closely related to the economy and human lifespan. It is the root cause of natural disasters such as floods and droughts that affect people all over the world every year. The accuracy of rainfall forecasts is critical for countries like India, whose economy is primarily based on agriculture. Because of the complex existence of the atmosphere, applied mathematics techniques struggle to provide reasonable precision for precipitation predictions. Regression can be used in the prediction of precipitation using machine learning techniques. The aim of this project is to provide non-experts with easy access to the techniques and approaches used in the field of precipitation prediction, as well as a comparative analysis of the various machine learning techniques.

***Keywords:*** Rainfall, Prediction, Machine Learning, Regression, Accuracy

## I. INTRODUCTION

Rainfall forecasting is critical since heavy and erratic rainfall can have a wide range of consequences, including crop failure and property damage. A better forecasting model is needed for early warning, which can reduce risks to life and property while also better managing agricultural farms. This forecast mostly benefits farmers, and water supplies can be used more effectively. Rainfall prediction is a difficult challenge, and the findings must be correct. There are numerous hardware devices available for predicting rainfall based on weather conditions such as temperature, humidity, and pressure. Since conventional approaches are inefficient, we can achieve reliable results through using machine learning techniques. We can simply do it by analysing historical rainfall data and forecasting rainfall for future seasons. We can apply many techniques, such as classification and regression, depending on the requirements, and we can also quantify the error between the real and predicted values, as well as the accuracy. Since various techniques yield varying degrees of accuracy, it is critical to choose the appropriate algorithm and model it in accordance with the requirements. Regression analysis is concerned with the dependency of one variable (referred to as the dependent variable) on one or more other variables (referred to as the independent variables), and is useful for estimating and/or projecting the mean or average value of the former in terms of known or fixed values of the latter. For example, a person's salary is determined by his or her experience. In this case, the experience factor is an independent variable, whereas the salary is a dependent variable. The relationship between a single dependent variable and a single independent variable is described by simple linear regression.

**The advantages of regression analysis are:**

1. It is a powerful technique for testing relationship between one dependent variable and many independent variables.
2. It allows researchers to control extraneous factors.
3. Regression asses the cumulative effect of multiple factors.
4. It also helps to attain the measure of error using the regression line as a base for estimations.

## II. RELATED WORK

Thirumalai, Chandrasegar, and colleagues [1] analyse the amount of rainfall in previous years based on crop seasons and forecast rainfall for future years. Rabi, Kharif, and Zaid are the crop seasons. For early prediction, the linear regression approach is used. Rabi and kharif were used as variables, and if one was supplied, the other could be predicted using linear regression. The standard deviation and mean were also determined in order to forecast crop seasons in the future. This implementation will be used to help farmers decide which crops to harvest based on crop seasons.

Geetha, A., and G. M. Nasira. [2] introduce a model that forecasts weather conditions such as flooding, fog, thunderstorms, and cyclones, allowing people to take preventive steps. To model the decision trees, data mining techniques were used, as well as a data mining method called Rapid miner. Trivandrum data set with attributes such as day, temperature, dew point, strain, and so on. The dataset is divided into two parts: training and research, and the decision tree algorithm is used. The precision is determined, and the real and expected values are compared. The accuracy is 80.67, and to achieve a high value, soft computing techniques such as fuzzy logic and genetic algorithms can be used.

Parmar, Aakash, Kinjal Mistree, and Mithila Sompura [3] address the various methods for predicting rainfall for weather forecasting and their limitations. Various neural networks algorithms used for prediction are discussed in depth, along with their measures. This categorises various methods and algorithms used for rainfall prediction by various researchers in today's period. Finally, the paper's thesis is presented. Context research on several machine learning models has been completed. ARIMA Model, Artificial Neural Network and its variants such as Back-Propagation Neural Network - Cascade Forward Back Propagation Network Layer Recurrent Network, Self-Organizing Map, and Support Vector Machine, Collected, surveyed, and table provides categorization of various rainfall prediction approaches.

Dash, Yajnaseni, Saroj K. Mishra, and Bijaya K. Panigrahi [4] used artificial intelligence techniques such as Artificial Neural Network (ANN), Extreme Learning Machine (ELM), and K nearest neighbour (KNN) to forecast summer and post-monsoon rainfall. The dataset used is the Indian Institute of Tropical Meteorology's time series data for Kerala from 1871 to 2016. (IITM). The data is then pre-processed and normalised before being split into training and testing groups. The data up to 2010 was used as the training collection, and the data from 2011 to 2016 was used as the test set. The algorithms listed above were used, and their output was calculated using MAE, RMSE, and MASE. In comparison to the others, the ELM algorithm generated more accurate results.

According to Singh, Gurpreet, and Deepak Kumar[5,] many machine learning algorithms are used for rainfall prediction, and they have used a hybrid method that combines two techniques, Random forest and Gradient boosting, with many machine learning techniques such as ada boost, K-Nearest Neighbor (KNN), Support vector machine (SVM), and Neural Network (NN).

These have been applied to North Carolina rainfall data from 2007 to 2017, and the output is measured using various metrics such as F-score, precision, accuracy, and recall. Finally, eight hybrid models were proposed, with the superior being Gradient boosting-Ada boost, which produced good performance.

Kar, Kaveri, Neelima Thakur, and Prerika Sanghvi [6] used fuzzy logic to forecast rainfall based on temperature data in a geographic area. The fuzzy model was employed. Since the forecast is not reliable due to other climatic factors, they have found other contributing factors such as humidity and studied the advantages of fuzzy method over other techniques.

Sardeshpande, Kaushik D., and Vijaya R. Thool [7] used artificial neural networks, back propagation (BPNN), radial base function (RBFNN), and generalised regression (GRNN) on rainfall data from India, primarily from the Nanded district in Maharashtra, and the data was normalised between 0 and 1, and the algorithms were applied, and their output was measured and compared. When compared to GRNN, BPNN and RBFNN produced better performance.

Chen, Binghong, et al. [8] focus on non-linear machine learning approaches including gradient boosting decision tree model and deep neural networks for short-term rainfall prediction, and these algorithms were built on Alibaba cloud and data was collected from different sites, and effectiveness is calculated using classification metrics AUC, F1 score, precision and accuracy, and Regressio. It was discovered that DNN performed better than ECData.

Moon, Seung-Hyun, and colleagues [9] introduce an early warning system (EWS) that generates a signal when it exceeds a threshold limit that indicates Warning before 3 hours. This was accomplished by the use of machine learning techniques. Data from South Korea was collected from 2007 to 2012, and output was calculated using various metrics, resulting in the development of a confusion matrix. It was suggested to use logistic regression with feature selection and PCA. The F-measure is used to estimate the model's performance.

## III. PROPOSED ARCHITECTURE

Precipitation is predicted using a statistical model. The first step is to translate data into the proper format for conducting experiments, followed by a thorough data analysis and observation of variance in rainfall patterns. We predict rainfall by dividing the dataset into training and testing sets, then applying various machine learning approaches (MLR, SVR, and so on) and statistical techniques to compare and draw conclusions about the various approaches used. We attempt to mitigate the error using a variety of approaches. Since the dataset is so huge, feature reduction is used to boost accuracy while also reducing computation time and storage. Principal Component Analysis (PCA) is a method for removing important variables from a large number of variables. It extracts low-dimensional sets with the aim of capturing as much information as possible. When there are few variables, visualisation becomes more important. It is accomplished by using a covariance matrix and extracting Eigen values from it. Using PCA, we reduced the attributes in our dataset by considering only rainfall data from three consecutive months and annual data from each subdivision. Techniques employed: Clearly, it is nothing more than an extension of simple regression toward the mean.

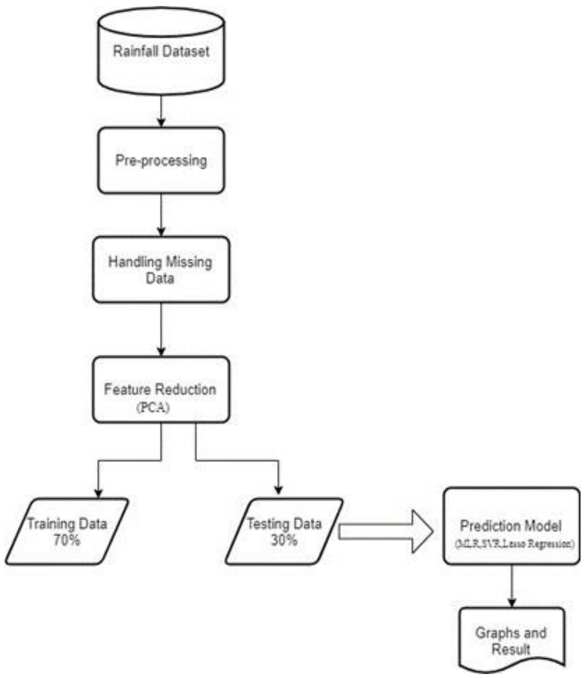
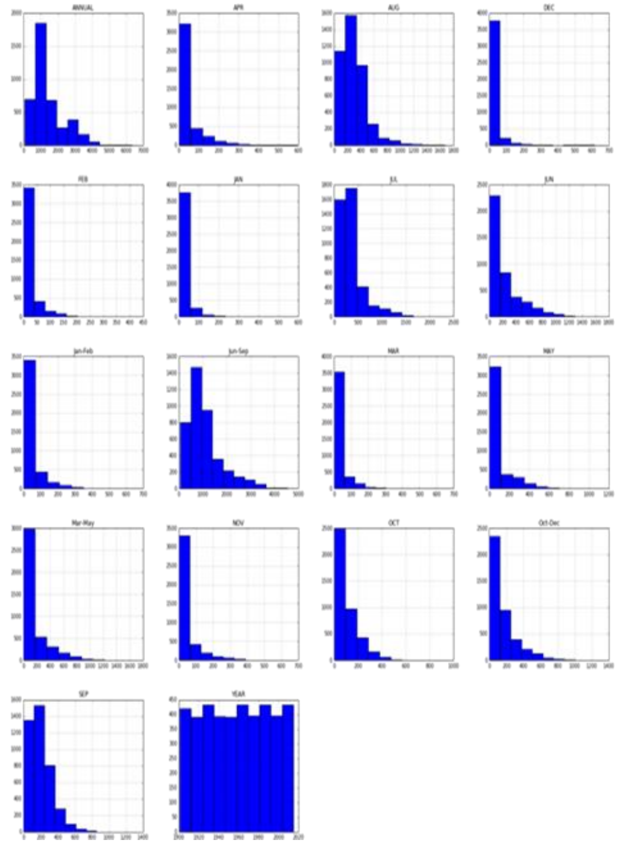



***Fig 1**. Rainfall Prediction Model*

Support Vector Regression Machine Learning and Data Science with the word SVM or support vector machine but SVR that is support vector regression is a bit different from SVM that is support vector machine as the name implies that is integration algorithm so we can use SVR for working with continuous value rather than classification which is SVM Support Vector Machines can perform both linear and nonlinear regression, which is referred to as Support Vector Regression. Support Vector Regression, rather than attempting to fit the largest possible street between two groups while restricting margin violations, seeks to fit as many instances as possible on the street while limiting margin violations. Epsilon, a hyperparameter, is used to calculate the lane dimension.

## IV. RESULTS AND OBSERVATION

Rainfall data from 1901 to 2015 is obtained, analysed, and plotted to better understand rainfall in different regions. The histograms plotted for monthly, annual, and three-month consecutive rainfall data are shown below. The amount of rainfall (Y-axis) increases during the months of July, August, and September.



## V. CONCLUSION

This project focused on rainfall estimation, and it is estimated that SVR is a valuable and adaptable technique, assisting the client in overcoming impediments related to distributional properties of fundamental variables, information geometry, and the common issue of model overfitting. The selection of bit capacity is critical for SVR show. We advise tenderfoots to use straight and RBF pieces for direct and non-straight relationships separately. As an expectation technique, we see that SVR outperforms MLR. MLR cannot detect non-linearity in a data set, so SVR is useful in such cases. We also process Mean Absolute Error (MAE) for both MLR and SVR models to evaluate model execution. Finally, we will look at the presentation of the SLR, SVR, and tuned SVR models. The tuned SVR model, as anticipated, provides the best prediction.

## VI. REFERENCES

[1] Thirumalai, Chandrasegar, et al. "Heuristic prediction of rainfall using machine learning techniques." 2017 International Conference on Trends in Electronics and Informatics (ICEI). IEEE,

2017.

[2] Geetha, A., and G. M. Nasira. "Data mining for meteorological applications: Decision trees for modeling rainfall prediction." 2014 IEEE International Conference on Computational Intelligence and Computing Research. IEEE, 2014

[3] Parmar, Aakash, Kinjal Mistree, and Mithila Sompura. "Machine learning techniques for rainfall prediction: A review." 2017 International Conference on Innovations in information Embedded and Communication Systems. 2017.

[4] Dash, Yajnaseni, Saroj K. Mishra, and Bijaya K. Panigrahi. "Rainfall prediction for the Kerala state of India using artificial intelligence approaches." Computers & Electrical Engineering 70 (2018): 66-73.

[5] Singh, Gurpreet, and Deepak Kumar. "Hybrid Prediction Models for Rainfall Forecasting." 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2019.

[6] Kar, Kaveri, Neelima Thakur, and Prerika Sanghvi. "Prediction of Rainfall Using Fuzzy Dataset." (2019).

[7] Sardeshpande, Kaushik D., and Vijaya R. Thool. "Rainfall Prediction: A Comparative Study of Neural Network Architectures." Emerging Technologies in Data Mining and Information Security. Springer, Singapore, 2019. 19-28.

[8] Chen, Binghong, et al. "Non-Linear Machine Learning Approach to Short-Term Precipitation Forecasting." (2018).

[9] Moon, Seung-Hyun, et al. "Application of machine learning to an early warning system for very short-term heavy rainfall.―Journal of hydrology 568 (2019): 1042-1054.

[10] https://data.gov.in/resources/subdivision-wise-rainfall-and- its-departure-1901-2015