

# Prediction of Rainfall Using Machine Learning Algorithms

Ishwarya G, Santhrupthi M B, Shanthi B, Asst. Prof. Varsha N

Dept. of Information Science and Engineering,  
Vidya Vikas Institute of Engineering and Technology,  
Mysuru, India.

ishwaryalote207@gmail.com, Santhuuppar123@gmail.com, shanthib98@gmail.com, mvarsha453@gmail.com

**Abstract-** Prediction of rainfall using precipitation forecast is significant as substantial precipitation can prompt numerous debacles. The expectation assists individuals with taking preventive measures and besides the forecast ought to be exact. There are two sorts of expectation transient precipitation forecast and long haul precipitation. Forecast generally transient expectation can give us the precise outcome. The fundamental test is to construct a model for long haul precipitation forecast. Substantial precipitation expectation could be a significant disadvantage for geology division since it is firmly connected with the economy and lifetime of human. It is a reason for cataclysmic events like flood and dry spell that square measure experienced by people across the world every year. Precision of precipitation articulation has decent significance for nations like India whose economy is fundamentally reliant upon horticulture. The powerful idea of air, applied math procedures neglect to give reasonable precision to precipitation proclamation. The forecast of precipitation utilizing AI methods may utilize relapse. Expectation of this undertaking is to offer non-specialists simple admittance to the methods, approaches used in the area of precipitation forecast and give a near report among the different AI strategies.

**Keywords-** Rainfall, Prediction, Machine Learning, Regression, Accuracy, SVR, Linear Regression.

## I. INTRODUCTION

Precipitation estimating is vital in light of the fact that substantial and unpredictable precipitation can have numerous effects like obliteration of harvests and ranches, harm of property so a superior gauging model is fundamental for an early notice that can limit dangers to life and property and furthermore dealing with the horticultural homesteads in better manner.

This expectation chiefly helps ranchers and furthermore water asset scan be used productively. Precipitation forecast is a difficult assignment and the outcomes ought to be precise. There are numerous equipment gadgets for anticipating precipitation by utilizing the climate conditions like temperature, mugginess, pressure. These conventional strategies can't work in a proficient manner so by utilizing AI methods we can create precise outcomes.

We can take care of business by having the recorded information examination of precipitation and can foresee the precipitation for future seasons. We can apply numerous methods like order, relapse as indicated by the prerequisites and furthermore we can ascertain the blunder between the real and expectation and furthermore the exactness. Various strategies produce various exactness's so pick the right calculation and model it as indicated by the necessities.

### 1. Regression Analysis:

Regression analysis deals with the dependence variable of one variable (called as dependent variable) on one or more other variables, (called as independent variables) which is useful for estimating and/ or predicting the average or mean value of the former in terms of known or fixed values of the latter. For example, the salary of a person is based on the experience members here; the experience attribute is independent variable salary is dependent variable.

Simple linear regression is defines the relationship between a single dependent variable and a single independent variable. The below equation is the general formula for regression.  $y = \beta_0 + \beta_1 x + \epsilon$  where  $\beta_0$  and  $\beta_1$  are parameters, and  $\epsilon$  is a probabilistic error term.

Regression analysis is a vital tool for analyzing and modeling information. It is used for prediction analysis that is forecasting of rainfall or weather, predicting trends in business, finance, and marketing. It can also be used for correction of errors and also provide quantitative support.

#### The advantages of regression analysis are:

- It is an incredible method for testing connection between one ward variable and numerous free factors.
- It allows researchers to control extraneous factors.
- Regression assesses the cumulative effect of multiple factors.

- It also helps to attain the measure of error using the regression line as a base for estimations.

## II. LITERATURE REVIEW

**Hirumalai, Chandrasegar, et al. [1]** examines the measure of precipitation in past years as indicated by the harvest seasons and predicts the precipitation for future years. The yield seasons are Rabi, Kharif and Zaid. Straight relapse strategy is applied for early expectation. Here, Rabi and kharif were taken as factors on the off chance that one variable was given; other can be anticipated utilizing direct relapse. Standard deviation and Mean was likewise determined for future expectation of yield seasons. This execution will be utilized for ranchers to have a thought of which yield to collect as indicated by crop seasons.

**Geetha A. and G. M. Nasira. [2]** Executes a model which predicts the climate conditions like precipitation, mist, rainstorms and tornadoes which will be useful to individuals to take preventive measures. Information mining procedures were utilized and an information mining instrument named Rapid digger was utilized to show the choice trees. The informational collection of Trivandrum with ascribes like day, temperature, dew point, pressure and so on The dataset is partitioned into preparing set and testing set and choice tree calculation is applied. The exactness is determined, genuine and anticipated qualities are thought about. The precision is 80.67 and to accomplish high worth it very well may be reached out by applying delicate processing strategies like fluffy rationale and hereditary calculations.

**Parmar, Aakash, Kinjal Mistree, and Mithila Sompura [3]** examine the various techniques utilized for precipitation expectation for climate gauging with their limitations. The various neural networks algorithm which are used for prediction are discussed with their steps in detail categorizes various approaches and algorithms used for rainfall prediction by various researchers in today's era. Finally, presents conclusion of paper. Done the background work as some models of machine learning ARIMA Model, Artificial neural network and types like Back- Propagation Neural Network - Cascade Forward Back Propagation Network Layer Recurrent Network, Support Vector Machine and Self-Organizing Map, surveyed, Collected, and table presents categorization of different approaches of rainfall prediction.

**Dash, Yajnasen I, Saroj K. Mishra, and Bijaya K. Panigrahi [4]** has used artificial intelligence techniques like K nearest neighbor (KNN), Extreme Learning Machine (ELM), Artificial Neural Network (ANN) are applied for prediction of post monsoon and monsoon summer rainfall. The dataset used is the time series data of Kerala from 1871 to 2016 taken the data from Indian Institute of Tropical Meteorology (IITM). The data is

normalization and pre- processed was performed on the next data, that data is divided into training and testing the data. We take the data up to 2010 as training set and the data from 2011- 2016 taken as test set. The above mentioned algorithms were applied and its performance was calculated by using MASE, RMSE, and MAE. The ELM algorithm has given a accurate results compared to the others.

**Singh, Deepak Kumar and Gurpreet [5]** states that there are many machine learning algorithms applied for the rainfall prediction and they have used a hybrid approach that is combining two techniques, Random forest and Gradient boosting with many machine learning techniques they used like ada boost, K-Nearest Neighbor(KNN), Neural Network(NN) and Support vector machine(SVM). These have been applied on the rainfall data to predict rainfall of North Carolina from 2007 – 2017 and also the performance is calculated by applying different metrics F-score, accuracy, recall, precision. Finally, they used eight hybrid models have been proposed and Gradient boosting-Ada boost has the superior which exhibited good results.

**Neelima Thakur, Kar, Kaveri and Prerika Sanghvi [6]** they used the fuzzy logic approach for the prediction of rainfall on the data of temperature in a geographic location in an area. The fuzzy model has been applied Due to other climatic factors the prediction is not accurate in rainfall prediction so they have considered other influencing factors like humidity also analyzed the advantages of fuzzy model over other techniques.

**Kaushik D, Sardesh Pande, and Vijaya R. Thool [7]** has used the back propagation neural networks (BPNN), artificial neural networks (ANN), radial basis function (RBFNN) and generalized regression (GRNN) on the rainfall data of India mainly Nanded district, Maharashtra was considered then the data is normalized between 0 to 1 and the algorithms are applied and the performance accrued data of those was calculated and compared. BBFNN and RPNN have given good results compared to GRNN.

**Binghong, Chen, et al. [8]** they focuses on the non-linear machine learning approaches like gradient boosting decision tree model, deep neural networks for a short term prediction of rainfall. These algorithms were built on Alibaba cloud then data was collected from different sites. Then effectiveness is calculated by using classification metrics F1 score, AUC, precision and accuracy and by Regression metric RMSE and correlation. It has observed that DNN showed better result than ECData.

**Seung-Hyun, Moon, et al [9]** implements an early warning system (EWS) that produces a signal when it reaches a threshold limit that gives warning before 3 hrs. This was done by using machine learning techniques. South Korea data from 2007 to 2012 was taken, then

performance is measured by some criteria and a confusion matrix was produced. The logistic regression with PCA was proposed and feature selection. F-measure is calculated for estimating the efficiency of rainfall prediction model.

### III. PROPOSED METHOD

The prescient model is utilized to expectation of the precipitation. The initial step is changing over information in to the right arrangement to direct tests then, at that point make a decent examination of information and notice variety in the examples of precipitation. We foresee the precipitation by isolating the data set into preparing set and testing set then we apply distinctive AI draws near (MLR, SVR, and so forth) and factual strategies and look at and draw investigation over different methodologies utilized. With the assistance of various methodologies we endeavour to limit the mistake.

#### 1. Data set Description:

The data set [10] comprises of the estimation of precipitation from year 1901-2015 for each state.

- Data comprises of 19 ascribes (singular months, yearly, and Mixes of 3 successive months) for 36 sub divisions.
- The information is accessible just from 1950 to 2015 for a Portion of the regions.
- The ascribes are the measure of precipitation estimated in Mm.

As the dataset is exceptionally huge, highlight decrease is done with the goal that it improves the exactness, diminishes the calculation time and furthermore stockpiling. Numerous Linear Regression: Multiple relapse attempts to show the association between two or extra factors and a reaction by fitting a condition to decided data. Plainly, it's anything but's an expansion of straight forward relapse toward the mean.

The overall type of multivariable straight relapse model is:  $y = \alpha + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \epsilon$  where  $y$  = subordinate variable and  $x_1, x_2, \dots, x_k$  are free variables,  $\alpha, \beta$  are coefficients. Different relapse will demonstrate extra confounded relationship that comes from various choices along they ought to be utilized in cases any place one unequivocal variable isn't sufficiently clear to plan the connection between the autonomous and furthermore the variable amount.

#### 2. Support Vector Regression:

Backing Vector relapse AI and information science with the term SVM or backing vector machine yet SVR that is support vector relapse is a digit not the same as SVM that is support vector machine as the name recommends that is joining calculation so we can utilize SVR for working with constant worth rather than order which is SVM Support Vector Machines support straight and nonlinear relapse that we can allude to as Support Vector Regression. Rather

than attempting to fit the biggest conceivable road between two classes while restricting edge infringement, Support Vector relapse attempts to fit whatever number occasions as could be expected under the circumstances in the city while restricting edge infringement. The size of the path is estimated by a hyper boundary Epsilon.

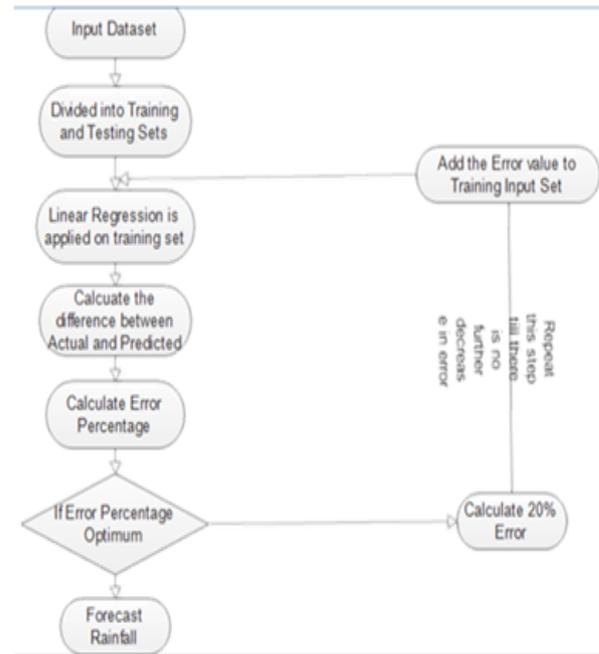


Fig 1. Rainfall Prediction Model.

#### 3. Kernel- The function is used to map a low dimensional data into higher dimensional data from the model.

- **Hyper plane**-in SVM this is a fundamentally The Separation line between the information classes likewise in SVR we will characterize it is as the line that will that will assist us with foreseeing the ceaseless worth or target esteem.
- **Limit line** - the SVM plane which makes envision the help vector can be on limit lines or outside the limit line isolates two classes in the idea same.
- **Vectors**-these are the information focuses which are nearest to the limit the distance of the fact of the matter is least.
- **SVR** performs straight relapse in higher dimensional space. We can consider SVR if every information point in the preparation addresses its own measurement. At the point when we assess piece between a test point and a point in the preparation set the subsequent worth gives you the organize of your test point in that measurement. The vector we get when we assess the test point for all focuses in the preparation set,  $k$  is the portrayal of the test point in the higher dimensional space. The condition of the hyper plane is  $Wx+b=0$  and the two conditions of limit lines is;

$$Wx+b=+e, Wx+b=-e$$

Equation that fulfill our SVR is  $e \leq y - Wx - b \leq +e$ .

SVR has an alternate relapse objective contrasted with straight relapse in direct relapse, we are attempting to limit the mistake between the forecast and information though in SVR an objective is to ensure that blunder don't surpass the edge.

#### 4. Algorithm:

- Rainfall prediction Input: Rainfall data set.
- **Output:** Accuracy or error of the prediction.
- **Step 1:** Import the rainfall data set csv file.
- **Step 2:** Fill the missing values with mean values of the data.
- **Step 3:** Scaling the data for features scaling the data to a fixed scale.
- **Step 4:** The data is divided into training set (70%) and testing set (30%).
- **Step 5:** Multiple Linear Regression algorithm, Support vector machine.

### IV. EXPERIMENTAL RESULTS

The information of precipitation from 1901-2015 is gathered an information is considered and plotted to comprehend the precipitation in different districts. The beneath is the histograms plotted for the precipitation information month to month, yearly and back to back of 90 days.

It is seen that there is emerge in volume of rainfall(Y-pivot) in the long periods of July, August and September. In the below graph it shows the accuracy value using an algorithm (SVM and LR). We use two algorithms to find the High Accuracy value to predict the rainfall prediction.

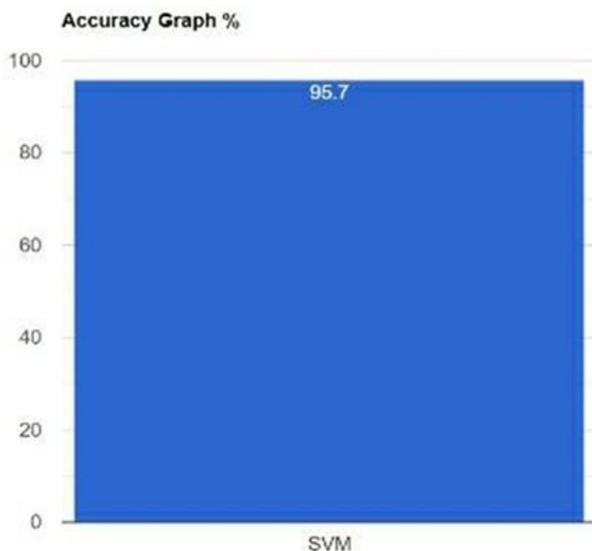


Fig 2. Accuracy Graph of Support Vector Machine.

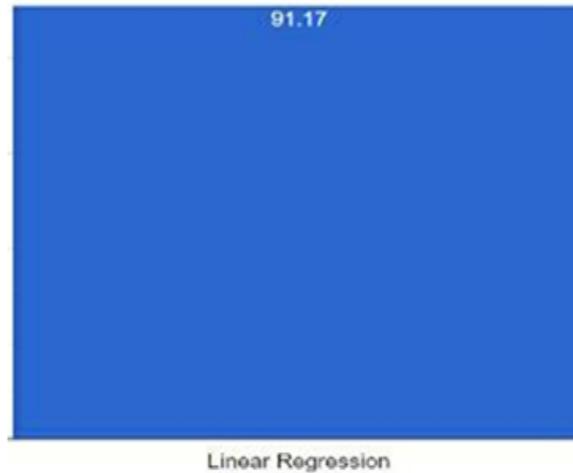


Fig 3. Accuracy Graph of Linear Regression Algorithm.

Table 1. Accuracy Value of Rainfall Prediction.

Algorithm to Prediction Model	Mean Absolute value
Linear Regression	91.17
Support VectorMachine	96.7

### V. CONCLUSION

This undertaking focused on assessment of precipitation and it is assessed that SVR is a significant and versatile technique, assisting the customer with dealing with the hindrances identifying with distributional properties of central components, calculation of the data and the typical issue of model over fitting.

The choice of touch limit is fundamental for SVR showing. We endorse novices to use straight and RBF piece for immediate and non-straight relationship exclusively. We see that SVR is superior to MLR as an assumption technique.

MLR can't get the non-linearity in an informational collection and SVR ends up supportive in such conditions. We furthermore measure Mean Absolute Error (MAE) for both SVR and MLR models to assess execution of the models. At last, we look at the presentation of SVR, SLR and tuned SVR model. True to form, the tuned SVR model gives the best expectation.

### ACKNOWLEDGMENT

We offer our thanks towards almighty for guiding us. We need to thank every single one of the people who helped us for the culmination of this paper. We are appreciative to our coordinators for all help and sponsorship to contact us. We thank all staff people from our college and our allies for their collaboration. Over all we need to thank our

people for their help, without which we would not have accomplished our goal.

## REFERENCES

- [1] Arun Kumar, Naveen Kumar, Vishal Vats, “Efficient Crop Yield Prediction using rainfall Using Machine Learning Algorithms”, International Research Journal of Engineering and Technology (IRJET)- e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume: 05 Issue: 06 | June-2018
- [2] Mrs. K.R. Sri Preethaa, S.Nishanthini, D.Santhiya, K.Vani Shree, “Crop Yield Prediction using rainfall”, International Journal On Engineering Technology and Sciences – IJETSTM ISSN(P): 2349-3968, ISSN (O):2349-3976 Volume III, Issue III, March- 2016.
- [3] J. Ramirez-Villegas and A. Challinor, “Assessing relevant climate data for agricultural applications”. *Agricultural & Forest Meteorology*, 2012, vol. 161(3), pp. 26–45.
- [4] Hernandez, Emilcy & Sanchez-Anguix, Víctor & Julián, Vicente & Palanca, J. & Duque, Néstor. (2016).”Rainfall Prediction: A Deep Learning Approach.” 151-162. 10.1007/978-3-319-32034-2\_13.
- [5] Zhou, Kanghui & Zheng, Yongguang & Li, Bo & Dong, Wansheng & Zhang, Xiaoling. (2019). Forecasting Different Types of Convective Weather: A Deep Learning Approach. *Journal of Meteorological Research*. 33.797-809. 10.1007/s13351-019-8162-6.
- [6] Senekane, Makhamsa & Mafu, Mhlambululi & Taele, Benedict. (2021). “Weather Nowcasting Using Deep Learning Techniques.” 10.5772/intechopen.84552.