# Utilizing Machine Learning Algorithms for Rainfall Analysis

Dr. Y. Jeevan Nagendra Kumar
Professor and Dean,
Department of Information Technology,
Gokaraju Rangaraju Institute Of Engineering and Technology,
JNTUH, Hyderabad, India
jeevannagendra@griet.ac.in

Shirisha. K
Information Technology,
Gokaraju Rangaraju Institute Of Engineering and Technology,
JNTUH, Hyderabad, India
shirishareddy1024@gmail.com

Niveditha. N
Information Technology,
Gokaraju Rangaraju Institute Of Engineering and Technology,
JNTUH, Hyderabad, India
bunnyniveditha28@gmail.com

Swapna. M
Information Technology,
Gokaraju Rangaraju Institute Of Engineering and Technology,
JNTUH, Hyderabad, India
swapnabai52@gmail.com

Pavitra Sagar
Information Technology,
Gokaraju Rangaraju Institute Of Engineering and Technology,
JNTUH, Hyderabad, India
sagarpavitra7@gmail.com

Prashanth. I
Information Technology,
Gokaraju Rangaraju Institute Of Engineering and Technology,
JNTUH, Hyderabad, India
inapanuriprasanth2002@gmail.com

**Abstract: Agriculture relies greatly on rainfall. Recent years witnessed a substantial improvement in the complexity of rainfall prediction. Rainfall forecast will provide valuable predictions to farmers to help them take the appropriate precautions to protect their crops from various weather conditions. Several techniques are available to predict rainfall. Machine Learning (ML) algorithms are particularly useful for predicting rainfall. As machine learning is a type of Artificial Intelligence (AI), it is essential for anticipating rainfall as it enables computer algorithms to make predictions more correctly without explicit guidance. Machine Learning (ML) uses previous data as input to predict the new output values. Meteorologists have attempted to predict future rainfall patterns via previous data. This method is referred to as rainfall forecasting. The primary objective of this research is to identify the best algorithm for predicting rainfall. In this work, SVR (Support Vector Regression) and linear regression strategies were used.**

**Keywords — Machine Learning, Linear Regression, SVR (Support Vector Regression), Supervised Learning, Classification, Regression,**

## I. INTRODUCTION

The weather forecasting data will be provided by the Indian Meteorological Department. This study concentrates on long-term rainfall estimates. The main purpose of this research work is to figure out how much rain will shower in a certain area or state in the future. The rainfall is simulated by using information gathered from the past. Understanding rainfall is absolutely vital everywhere in the universe and crucial to humankind. The tough function of methodically examining rainfall frequency lies in the meteorological agency. The varying atmospheric conditions make it difficult to forecast rainfall with any degree of precision. Rainfall is thought to be dependable during both the the wet and summer months. This is the first main reason, which needs a review of algorithms used to forecast rainfall. Different machine learning approaches, which include linear regression and SVR, are contrasted in order to find the best accurate model. Here, the UCI (University of California Irvine) repository's rainfall dataset is utilized. The numerous classification methods are explored and contrasted in this paper. The dataset is non-linear, and Support Vector Regression (SVR) exceeds regular Regression, thus we may state so. Various machine learning approaches, like linear regression and SVM classifiers are examined to identify the most accurate model. This article will compare and analyze different classification methods that are currently available.

## II. LITERATURE SURVEY

[1] Geetha worked on predicting the rainfall using Artificial Neural Network (ANN) algorithm. A module has been created to predict monthly rainfall over the Chennai region using the attributes like maximum and minimum temperature, relative humidity,

978-1-6654-6487-1/23/$31.00 ©2023 IEEE
DOI 10.1109/ICSMDI57622.2023.00069

speed of wind, and direction of wind.
[2] Zahoor jan described a rainfall prediction system that uses the deep learning K-nearest neighbor (KNN) technique. Using a given single K value, a total number of nearest neighbors are found that are useful in determining the class label for unknown data. He mentioned that similar parameters are clustered into the same kind of cluster when a dataset's class or category is determined using KNN. He described that this algorithm is not needed much time for train fin of classification and regression while it may cause inaccurate results when used an incorrect k value.

[3] Somvanshi used the ANN model for predicting the rainfall in Hyderabad by using the past four months' rainfall as the inputs. He also compared ANN with the ARIMA techniques. S. Chattopadhyay and M. Chattopadhyay employed two criteria for rainfall forecasting: minimum temperature and maximum temperature. For training the data they used the learning algorithms Levenberg-Marquardt (LM) and Conjugate Gradient Descent (CGD).

[4] S.K. Nanda,et.al (2013)[5] and the others used ARIMA, a sophisticated statistical model, and MLP (Multi Layer Perceptron), LPE (Legendre Polynomial Equation), and FLANN, three ANN models, to forecast rainfall (Functional-Link Artificial Neural Network).
They compared FLANN to the ARIMA model and stated that the FLANN model provides predictions with a higher degree of accuracy. They used the model called ANN for rainfall prediction and the parameters chosen for the model are temperature from min-max, mean sea level pressure, speed of wind, and rainfall.

[5] Hanna Mayer proposed a paper that studied the retrieval of optical rainfall using four machine learning approaches, and found that there was no change in the method results. and found that the average neural network performed the best, as did neural networks in general. He rated Neural networks as the most suitable algorithm because of its low computation times.

[6] Emilcy Hernandez and the others built a Deep Learning-based architecture to forecast the following day's quantity of daily precipitation. It has a perceptron with multilayers for the prediction task as well as an autoencoder for minimizing and capturing non-linear correlations between characteristics. They Compared this design to other earlier suggestions, and proved that this architecture shows an improvement in the capability to forecast the cumulative daily perception of the following day.

## III. EXISTING WORK

The existing research work have used various algorithms such as KNN, ANN. By using those algorithms, rainfall analysis has been carried out. But now in the proposed work, linear regression and SVR model are used for analyzing the amount of rainfall.

## IV. METHODOLOGY

Prediction of rainfall is nowadays very crucial because by that we can plan ourselves for our comfort. For farming also, we need water and this prediction helps the farmers a lot. The methods which we have used here are Linear Regression and SVR [8]. These two algorithms results are compared and then we choose the best algorithm which suits this prediction. Using these algorithms, we calculate absolute mean errors and the algorithm which gives less error gives better results [9]

### A. Data Collection

The source data for this investigation were gathered from an Indian dataset and data is gathered by many surveys and observations. With data, there are subdivisions and traits (individual months, annual, combinations of 3 consecutive months)[10]. Several of the subdivisions have data from 1950 to 2015 accessible. All the attributes have the sum of amount of rainfall in mm. The statistics was then transferred into a tabular Microsoft Excel file. The months were arranged in the table's column's row of tables comprising environmental variables [11]

### B. Data Preprocessing

The info Transformation, missing total quality management, and divided into training and testing all seem to be parts of preprocessing. Data from an Indian set of data is generated

during the period of 1901 to 2015. We must erase every data points since the data is incomplete [12].

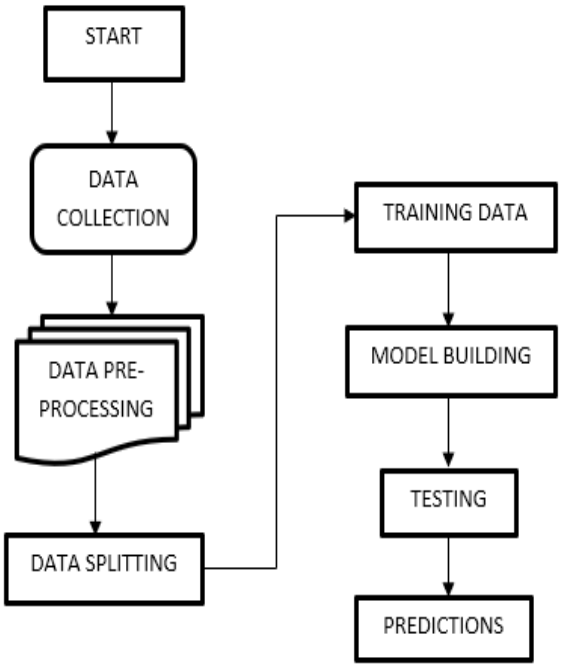


Fig 1: Workflow of proposed system

## C. Linear Regression

The earliest algorithm for machine learning is linear regression. This algorithm is mostly used for predictive analysis [13] For numerical variables or similar continuous variables like pay, age, etc., linear regression provides predictions. The link between dependent and independent variables is what linear regression is all about [14] Given its properties, linear regression may be applied to investigate the correlation between the measured values of the dependent variable and the value of the independent variable. The link between the variables is shown by the linear regression model as a sloping straight line[15]

## D. Support Vector Regression (SVR)

Support Vector Regression refers to supervised ml algorithms and relevant teaching algorithms that assess data for regression analyses and classification [16] Support Vector Machine, or SVM, is the framework around which SVR is built. One of the most formidable models that could be applied to classification issues or processing information when it cannot be separated linearly [17] SVR is a hyperplane drawn between the boundary points.

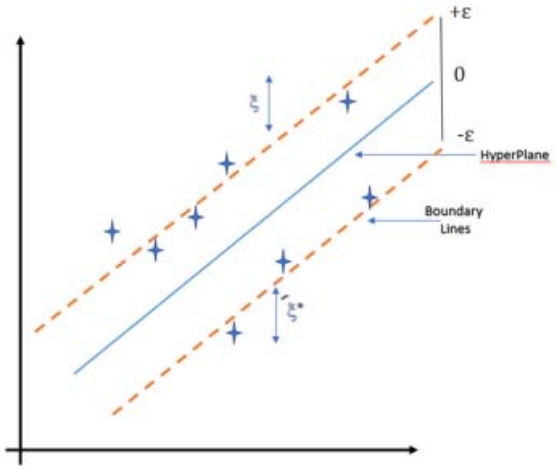


Fig 2: SVR

## E. Procedure

This dataset contains data on the monthly rainfall for 36 Indian meteorological subdivisions. The finding corresponds to specifics on the district-level rainfall average (in millimetres) calculated utilizing data for the years 1951 to 2000.Here Dataset is divided into two divisions namely dataset1 and dataset2. The dataset1 has average rainfall from 1951-2000 for each district, for every month dataset2 shows the average annual precipitation data for each state from 1901 to 2015 [18]

putting data into the right format because then studies may be performed out. Evaluate the information thoroughly and note variations in the rainfall pattern. [19]

Finally, by separating the information into training and validation, we attempt to calculate the average rainfall. In order to predict, we use a variety of machine-learning and statistical techniques (SVR, etc.), and we investigate these techniques. We strive to reduce inaccuracy by utilizing a wide range of tactics.

In these we calculate mean absolute error using algorithms called Linear regression and SVR compared with SVR, Linear regression gives better results in rainfall analysis.

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

It is average of absolute of actual minus forecast to actual.
The dataset is taken from the Kaggle source and utilised. To train the model 80% was used,while the 20% which is remained was used to test the results and train the model.

The rainfall prediction using Machine learning includes the main sections as data for input, pre-processing the data splitting of data. At the starting stage of the project, we will load the dataset into the system.

To adequately process the dataset and extract its attributes, pre-processing is needed. With this, the dataset may be divided into training and testing halves.

By applying Linear regression and SVR algorithms. The expected result should then be compared to the initial data set. It was afterwards assessed how accurate the model was using test samples.

Furthermore, we compared the model's accuracy with these two algorithms. Then ultimately concluded that the Linear regression algorithm increases the accuracy of the model. As a result, we trained the model using a Linear regression approach.

## V. RESULT

Data visualization is the depiction of comprehending data by displaying it in a graphical environment in order to find and disclose the designs, inclinations, and relationships.

The following are a few popular plotting collections:

- MatplotlibSmall level, gives users a lot of freedom.
- Pandas Visualization feels at secure using this limit. It may be created with Matplotlib.
- Plotly: Can create interactive plots for visualization.

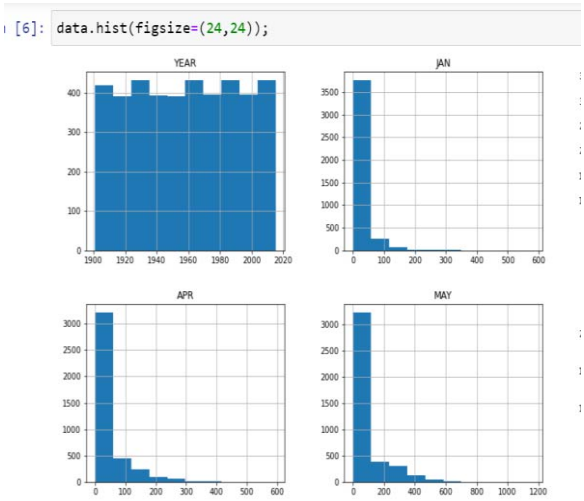


Fig 3: Histograms

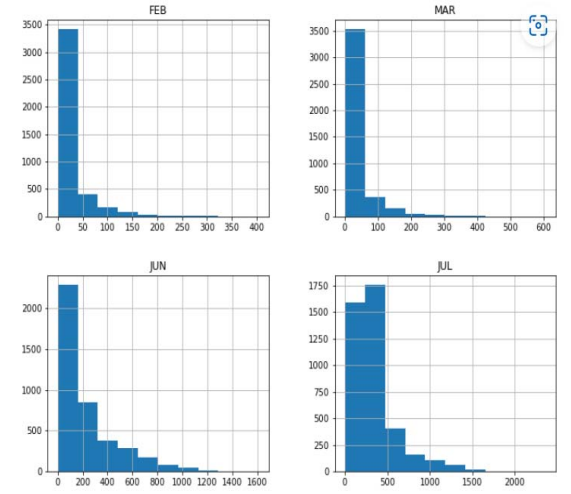


Fig 4: Histograms

As we know that amount of rainfall is calculated in mm. These histograms say the amount of rainfall for the whole year and also for all the months. So, by looking at those histograms we could predict the amount of rainfall every month.

We may determine how closely or distantly the qualities are connected to one another using correlation. The number 1 denotes a positive correlation between the variables. The dark colour of the value 0 indicates a stronger negative association between the qualities.

```
12]: plt.figure(figsize=(11,4))
     sns.heatmap(data[['Jan-Feb','Mar-May','Jun-Sep','Oct-Dec','ANNUAL']].corr(),annot=True)
     plt.show()
```


Fig 5: Correlation Diagram

Correlation matrix is a set of variables that are plotted against one other and it depicts the correlation between all the possible pairs in the matrix. Here we have used all the names of the months like january, febraury, march, april, may, june, july, august etc.

And also, we took column names as January-febraury, march-may, june-sep, oct-dec. We have also included Annual as one column.

```
19]: from sklearn import linear_model

     # Linear model
     reg = linear_model.ElasticNet(alpha=0.5)
     reg.fit(X_train, y_train)
     y_pred = reg.predict(X_test)
     print (mean_absolute_error(y_test, y_pred))

96.32435229744083
```

Fig 6: Linear Model

```
In [22]: from sklearn.svm import SVR

         # SVM model
         clf = SVR(gamma='auto', C=0.1, epsilon=0.2)
         clf.fit(X_train, y_train)
         y_pred = clf.predict(X_test)
         print (mean_absolute_error(y_test, y_pred))

127.1600615632603
```

Fig 7: SVR Model

These are the Prediction snapshots of rainfall. SVR is built from the model called SVM, so we import the svr algorithm from SVM.

Here we found absolute mean errors for both the algorithms. As we see, the Linear model gives 96.3 and SVR model gives the error 127.1.

So, through the Linear Regression algorithm we got less error so we say that Linear model gives better rainfall analysis than SVR model.

## VI. CONCLUSION

The two strategies for the model evaluation were utilized in this work. The algorithms we used were Linear Regression and SVR (Support Vector Regression). By using these algorithms, we have calculated the absolute mean error for both of them in which Linear Regression gave us the best results.
These results highlight the characteristics of rainfall that, in an ideal world, climate models should be able to simulate (or that should be bias-corrected). If seasonal climate estimates are used to inform crop models. The rise in model score here when reanalysis rainfall is gradually changed shows that improvements in simulations are projected to lead to more precise yield estimates..

## VII. FUTURE ENHANCEMENTS

In this paper, we can use other algorithms which can produce better results. And also we can use other datasets with other algorithms. You can use the Ann algorithm also to get the results.

## REFERENCES


[1] Geetha, G., and R. Samuel Selvaraj."Prediction of monthly rainfall in chennai using back propagation neural network model."International Journal of Engineering Science and Technology 3.1 (2011).

[2] Jan, Zahoor, et al. "Seasonal to interannual climate prediction using data mining KNN Conference. Springer, Berlin, Heidelberg, 2008.

[3] Somvanshi, V. K., et al. "Modeling and prediction of rainfall using artificial neural technique." International Multi Topic networks and ARIMA techniques." J. Ind. Geophys. Union 10.2 (2006): 141-151.

[4] Nanda, Santosh Kumar, et al. "Prediction of rainfall in India using Artificial Neural Network (ANN) models." International Journal of Intelligent Systems and Applications 5.12 (2013): 1.

[5] Meyer, Hanna, et al. "Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals." Atmospheric research 169 (2016): 424-433.

[6] Hernández, Emilcy, et al. "Rainfall prediction: A deep learning approach." International conference on hybrid artificial intelligence systems. Springer, Cham, 2016.

[7] Hanna Meyer, Meike Kuhnlein, Tim Applehans, Thomas Nauss,"Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrieval" in the year 2016

[8] Gupta D, Ghose U. A Comparative Study of Classification

Algorithms for Forecasting Rainfall. IEEE. 2015.
[9] Wang J, Su X. An improved K-Means clustering algorithm. IEEE. 2014.
[10] Atef Zaguia, Vikram Raju, Y. Jeevan Nagendra Kumar, Umashankar Rawat,"Secure Vertical Handover to NEMO using Hybrid Cryptosystem" Hindawi, Article ID 6751423, Hindawi
[11] Y. Jeevan Nagendra Kumar, Dr. T. V. Rajini Kanth, "GIS-MAP Based Spatial Analysis of Rainfall Data of Andhra Pradesh and Telangana States Using R", International Journal of Electrical and Computer Engineering (IJECE), Vol 7, No 1, February 2017, Scopus Indexed Journal, ISSN: 2088-8708
[12] B Sankara Babu, A Suneetha, G. Charles Babu, Y. Jeevan Nagendra Kumar, G. Karuna, "Medical Disease Prediction using Grey Wolf optimization and Auto Encoder based Recurrent Neural Network" Periodicals of Engineering and Natural Sciences, Vol 6 Issue 1 Pg: 229-240 ISSN 2303-4521 Jun 2018
[13] Dr. Y. Jeevan Nagendra Kumar, Guntreddi Sai Kiran, Partapu Preetham, Chila Lohith, Guntha Sai Roshik, G. Vijendar Reddy, "A Data Science View on Effects of Agriculture & Industry Sector on the GDP of India" International Journal of Recent Technology and Engineering, Volume-8, Issue-1, May 2019, ISBN: 2277-3878.
[14] Y. Jeevan Nagendra Kumar, N. Kameswari Shalini, P.K. Abhilash, K. Sandeep, D. Indira, "Prediction of Diabetes using Machine Learning" International Journal of Innovative Technology and Exploring Engineering, ISSN: 2278-3075, Volume-8 Issue-7 May 2019
[15] Y. Jeevan Nagendra Kumar, B. Mani Sai, Varagiri Shailaja, Singanamalli Renuka, Bharathi Panduri, "Python NLTK Sentiment Inspection using Naïve Bayes Classifier" International Journal of Recent Technology and Engineering, ISSN: 2277-3878, Volume-8, Issue-2S11, Sep 2019
[16] Srikanth Bethu, V Sowmya, B Sankara Babu, G Charles Babu, Y. Jeevan Nagendra Kumar, "Data Science: Identifying influencers in Social Networks", Periodicals of Engineering and Natural Sciences, ISSN 2303-4521 Vol.6, No.1, pp. 215~228
[17] Y Jeevan Nagendra Kumar, V Spandana, VS Vaishnavi, K Neha, VGRR Devi, "Supervised Machine Learning approach for Crop Prediction in Agriculture Sector", IEEE - 5th International Conference on Communication and Electronics Systems (ICCES), ISBN: 978-1-7281-5370-4 pg: 736-741
[18] D. Srinivasa Rao, Ch. Ramesh Babu, Y. J. Nagendra Kumar, N. Rajasekhar, T. Ravi, "Medical Image Fusion Using Transform Based Fusion Techniques", International Journal of Recent Technology and Engineering, Volume-8 Issue-2 ISSN: 2277-3878
[19] Atef Zaguia, Vikram Raju, Y. Jeevan Nagendra Kumar, Umashankar Rawat,"Secure Vertical Handover to NEMO using Hybrid Cryptosystem" Hindawi, Article ID 6751423, Hindawi.