www.revistageintec.net
ISSN: 2237-0722
revista GEINTEC
Gestão, Inovação e Tecnologias

# Machine Learning Models Applied for Rainfall Prediction

Gujanatti Rudrappa[1]; Nataraj Vijapur[2]; Rajesh Pattar[3]; Ravi Rathod[4]; Rashmi Kulkarni[5]; Vudu Sree Chandana[6]; Sateesh N. Hosmane[7]

[1]Department of ECE, KLE Dr. MSSCET, *(VTU)*, Belagavi, India.
[1]rudraguj@gmail.com
[2]Department of ECE, RVITM, Bangalore, India.
[2]nvijapur@gmail.com
[3]Department of ECE, KLE Dr. MSSCET, (VTU), Belagavi, India.
[3]rajeshpattar26@gmail.com
[4]Department of ECE, KLE Dr. MSSCET, (VTU), Belagavi, India.
[4]ravirathod3105@gmail.com
[5]Department of ECE, KLE Dr. MSSCET, (VTU), Belagavi, India.
[5]rashmikotesh@gmail.com
[6]Department of ECE, KLE Dr. MSSCET, (VTU), Belagavi, India.
[6]Vs.chandana0702@gmail.com
[7]Department of Chemical Engineering, KLE Dr. MSSCET, (VTU), Belagavi, India.
[7]satishosmane@gmail.com@gmail.com

**Abstract**

*Predicting rainfall is an important step in generating data for climate impact studies. Rainfall predictions are a key process for providing climate impact assessments with inputs. A consistent rainfall pattern is typically good for normal plants; nevertheless, too much or too little rainfall can be disastrous to crops, even deadly. Drought can damage plants and lead to erosion, while heavy rainfall can encourage the growth of destructive fungi. Machine Learning (ML) can be helpful in overcoming such issues; for example, ML can be used to predict rainfall and apply it to foresee crop health and yield. Predictive analysis is a subset of data mining that forecasts future probabilities and patterns. Various sectors like the Agricultural Produce Markets Committee (APMC), Kisaan call centre, etc., can use proposed method, enabling the sector and farmers to obtain information on future precipitation, crop yields and crop health.*

**Key-words:** Rainfall, Rainfall Prediction, Machine Learning, Predictive Analysis.

## 1. Introduction

In the current situation, rainfall is regarded as the one of the sole causes of most important traits. In India farming is seen as one of the key factors in deciding the country's economy, and agriculture depends solely on precipitation. The prediction of rainfall is important since heavy or irregular precipitation leads to crop devastation and destruction of property. The proposed prediction method and the prediction for rainfall outcomes can help farmers increase crop yield and detect crop health. It can also assist farmers in making more efficient use of water resources, resulting in increased crop productivity and crop health. Apart from that, knowing the amount of rainfall in a certain location is critical in coastal locations all over the world. To establish a rainwater harvester in some of the places where there is a water scarcity, rainfall forecasting should be done ahead of time. Predicting rainfall is a difficult task, and the findings must be correct. The goal of this project is to use Machine Learning (ML) to predict rainfall. The proposed project will focus on a comparison of different machine learning (ML) algorithms for rainfall prediction. Because various methods provide various levels of accuracy, it is critical to select the appropriate method and model it according to the needs for reliable rainfall forecast. Pre-processing of the data obtained for rainfall prediction is necessary, which includes data cleaning and normalizing of weather factors. The processed data (meteorological parameters) will then be used to predict rainfall. Prediction results can then be applied to various aspects of agriculture. The agricultural aspects that can be derived from the prediction results include crop yield estimation under current rainfall predictions, crop health identification in the future, and soil conditions for the predicted rainfall.

## 2. Literature Survey

Kaushik Datta and Goutaman P [1] proposed how to anticipate rainfall using ML and Neural Networks (NNs), taking into account variables such as air temperature, humidity, wind speed, and sunshine duration. Rainfall prediction using linear regression was proposed by Moulana Mohammed et al. [2], which took into account weather characteristics such as air temperature, relative humidity, and wind speed for rainfall prediction. G. Bala Sai Tarun et al. [3] presented an approach to use “Artificail Neural Network (ANN)”, “Support Vector Machine (SVM)”, and “Logistic Regression” to predict rainfall with factors include temperature, humidity, and pressure parameters. G. Rudrappa [4, 5] proposed the used of k-means cluster and Content Based Image Retrieval method for classification of clouds which is considered as a weather parameter for rainfall prediction in the future scope of

proposed work. Girish L et al. [6] proposed several ML techniques such as “linear regression”, “SVM”, “k-Nearest Neighbour (KNN)” approach, and “decision tree algorithm” to make rainfall predictions. In comparison to all other algorithms, the SVM algorithm produced the maximum efficiency of 93 percent (i.e., F1-score of about 0.9). Aakash Parmar et al. [7] presented an overview of various ML algorithms for rainfall prediction, as well as issues that may arise when using different approaches for rainfall prediction. Prabakaran S et al. [8] focuses on the linear regression method for determining a relationship between dependent and nondependent variables. This method can provide a good estimate of rainfall for a given time period. It is concerned with gathering a data set and preparing it for future processing. The obtained data is further processed, and the outcome is projected. Box G et al. [9] proposed the use of “time series forecasting” for various applications, “ARMAX” is one of the most common methods used for the purpose. The “Auto Correlation Function (ACF)” and “Partial Auto Correlation Function (PACF)” are used to find a model that appears to match the series' behavior for further analysis and parameter estimation. The “SARIMA (Seasonal Auto Regression Integrated Moving Average)” model, which was tested in Sudan, was reported in this by Etuk et al. [10].

## 3. Methodology

The proposed work will use Machine Learning Techniques to predict rainfall. The data mining method aids in the discovery of a hidden pattern related with the target function (here it will be rainfall prediction). This method uses all of the variables that affect rainfall, such as humidity, pressure, wind direction, and wind speed, to forecast future rainfall. Prediction can be done using both supervised and unsupervised techniques. Past meteorological data, which can be viewed online, is used to identify the primary components that influence rainfall. These variables are then standardized and pre-processed. These normalized weather parameters can then be used to predict rainfall. This prediction outcome is then used in the agricultural system to estimate crop health and yield, which is the future scope of this proposal.

A survey on the various methods used for rainfall prediction provides information on the various methodologies used for rainfall prediction. Multi-layer Perception (MLP), Support Vector Machine (SVM), Back Propagation Neural Network (BPNN), and other techniques are used in these methods.

Fig. 1 - Block diagram to Predict Rainfall


**Weather Parameters**

Weather parameters include the parameters that are either collected for a specific region or obtained from secondary data which involves already collected weather parameters. The weather parameters may include air temperature, atmospheric pressure, humidity, precipitation, sunshine, etc. In present work we make use of dataset made available in [10]. It consisted of 109941 entries for 24 parameters. Cloud characteristics is also a parameter considered in this dataset.

**Pre-Processing**

In ML, data pre-processing is an important step that helps improve the quality of data and promotes the extraction of meaningful insights from the data. In Machine Learning, data pre-processing refers to the technique of preparing (cleaning and organising) raw data in order to build and train ML models. Data pre-processing converts raw data into a readable and understandable format. We require data pre-processing because it improves the accuracy, consistency, and completeness of our dataset. We also deal with the missing data if any by imputing the data wherever necessary. Also, we identify the outliers with the help of Inter Quartile Range (IQR) and remove those outliers so as to avoid over fitting.

**Normalization**

It's a technique that's frequently used in data preparation for ML. The importance of normalisation is to convert the values of numeric columns in the dataset to a common scale without distorting the range of values or losing data. Prior to using weather parameters in the ML model for rainfall prediction, those parameters must be normalised. The primary motivation for normalising data is to reduce the number of duplicates in the dataset.

**ML Model**

After normalising all of the data, we use Multi-Layer Perceptron (MLP) to predict rainfall, and we calculate the F1-score, recall, accuracy to ensure that our ML model is accurate. A MLP is a NN that connects multiple layers in a directed graph, which implies that the signal only goes one direction through the nodes. Aside from the input nodes, each node has a "nonlinear activation function". Backpropagation is a supervised learning technique used by an MLP. A feedforward ANN with multiple layers is known as MLP. There are at least three levels of nodes in an MLP: an input layer, a hidden layer, and an output layer. We also test out different ML models for rainfall prediction. Models such as Random Forest, Neural Network, Decision Tree, XGBoost, Catboost, and LightBGM are tested for rainfall prediction using the dataset mentioned above.

**Rainfall Prediction Result**

In this result we are predicting whether there will be rainfall next day or not is predicted. For this prediction we test various ML models such as Random Forest, Neural Network, Decision Tree, XGBoost, Catboost, and LightBGM.

## 4. Results

In the dataset [11] we have 24 parameters and we have 109941 values associated with these parameters. The dataset contains weather parameters Minimum and Maximum Temperature, Wind Speed, Wind Direction, Precipitation, Cloud, etc. The dataset contains entry from 2008 to 2017. Prior to selecting the features for rainfall prediction, we normalize the data so as to overcome the fitting problem. We divide the dataset into training and testing data with 75% for training and 25% for testing the different model prediction. Table 1 presents the details about the performance of different ML models.

In the results presented accuracy, precision, recall and F1 scores are estimated for different ML models. Precision refers to how similar measures are to one another; in this case, it refers to the percentage of our results that are relevant. Recall refers to the percentage of total relevant results correctly classified by the ML models. F1-score is the harmonic mean of the precision and recall of the model. A good F1-score is one with few false positives and few false negatives. The F1-score is assumed to be perfect if it is close to 1. The model fails when the F1-score reaches 0. From the results

obtained we could make out that XGBoost has better performance in terms of accuracy, precision, recall and F1 score. F1 score for XGBoost is highest as compared to other ML models.

Fig. 2 - Accuracy of ML Models

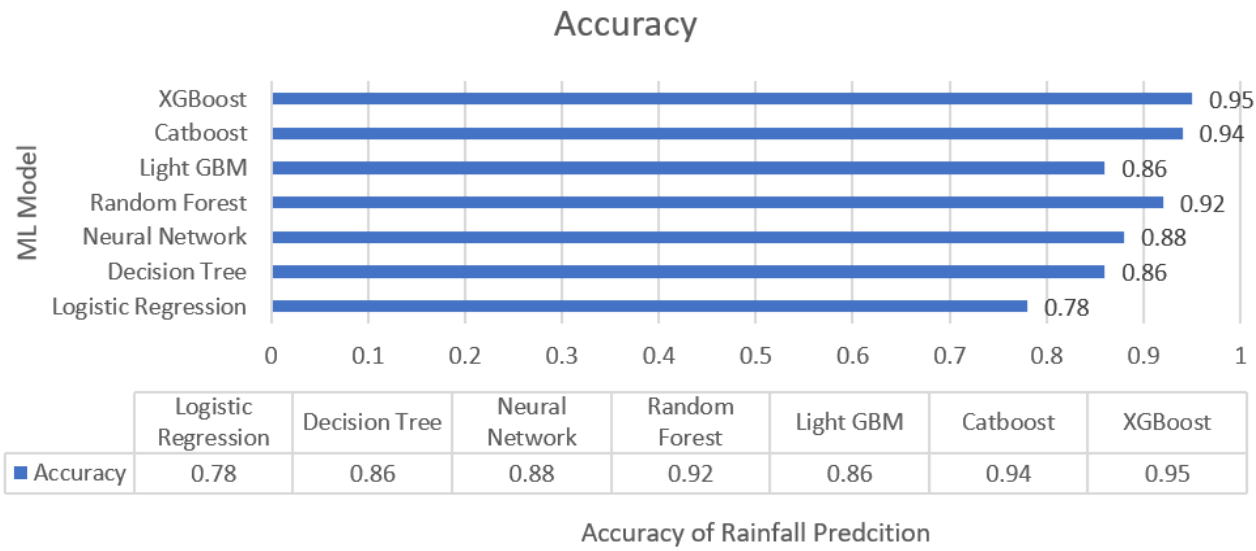
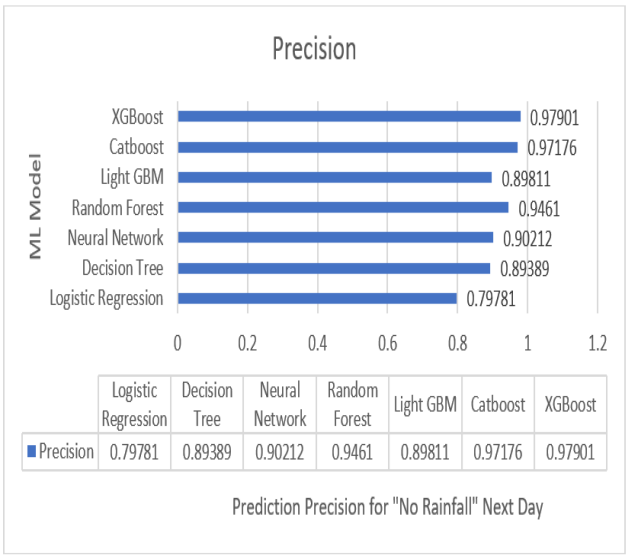



Fig. 2 above shows the accuracy of different ML models.

Table 1 - Comparison of different ML Model Performances

| ML Model | Rainfall Next Day | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| Logistic Regression | No | 0.78 | 0.79781 | 0.85698 | 0.82634 |
| | Yes | 0.78 | 0.74938 | 0.66319 | 0.70366 |
| Decision Tree | No | 0.86 | 0.89389 | 0.87048 | 0.88203 |
| | Yes | 0.86 | 0.80698 | 0.83975 | 0.82304 |
| Neural Network | No | 0.88 | 0.90212 | 0.90041 | 0.90126 |
| | Yes | 0.88 | 0.84601 | 0.84850 | 0.84725 |
| Random Forest | No | 0.92 | 0.94610 | 0.93163 | 0.93881 |
| | Yes | 0.92 | 0.89644 | 0.91769 | 0.90694 |
| Light GBM | No | 0.86 | 0.89811 | 0.87742 | 0.88765 |
| | Yes | 0.86 | 0.81646 | 0.84564 | 0.83079 |
| Catboost | No | 0.94 | 0.97176 | 0.92719 | 0.94895 |
| | Yes | 0.94 | 0.89458 | 0.95821 | 0.92531 |
| XGBoost | No | 0.95 | 0.97901 | 0.93321 | 0.95556 |
| | Yes | 0.95 | 0.90343 | 0.96898 | 0.93505 |

Fig. 3 - Prediction Precision for different ML Models



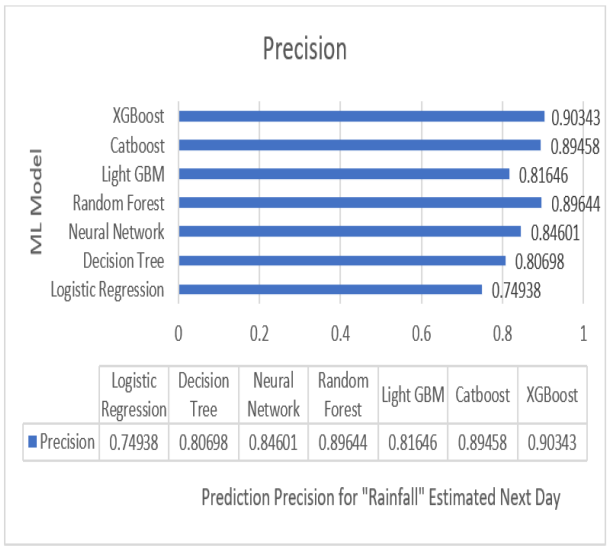
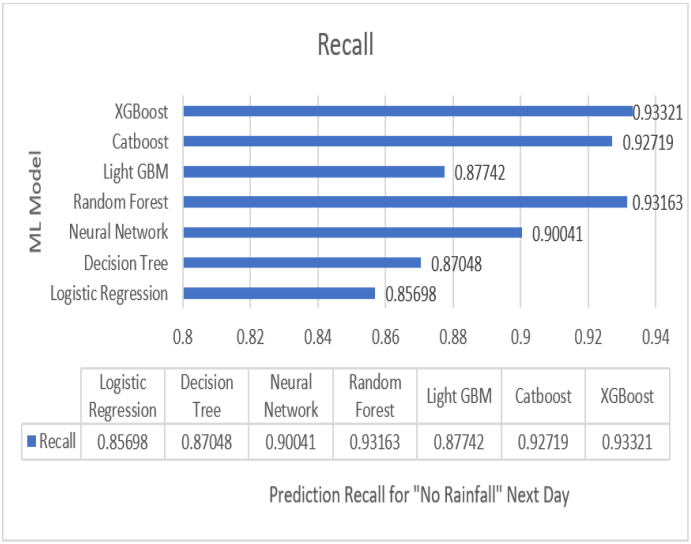




Fig. 3 shows the prediction precision for “No Rainfall” (no rainfall is predicted) and “Rainfall” (rainfall os predicted) estimated next day. Fig. 4 below shows the prediction Recall for “No Rainfall” (no rainfall is predicted) and “Rainfall” (rainfall os predicted) estimated next day. Fig. 5 below shows the prediction F1-Score for “No Rainfall” (no rainfall is predicted) and “Rainfall” (rainfall is predicted) estimated next day.

Fig. 4 - Prediction Recall for different ML Models



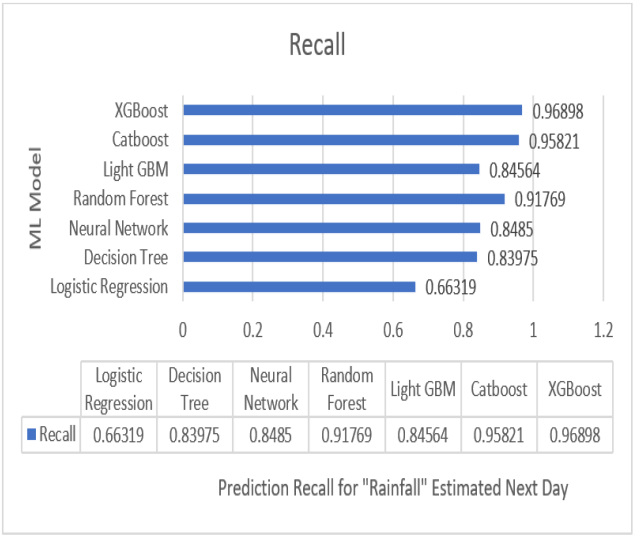
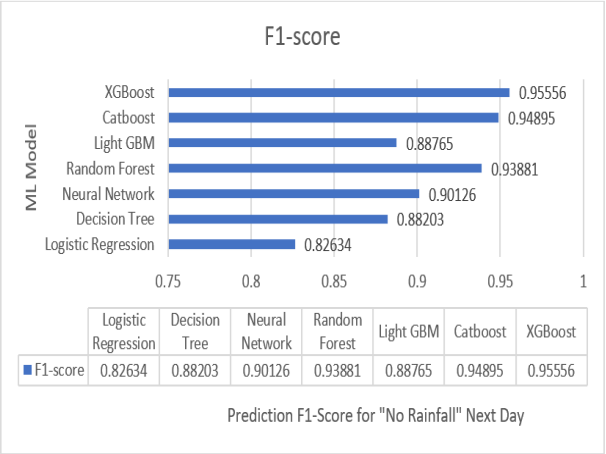
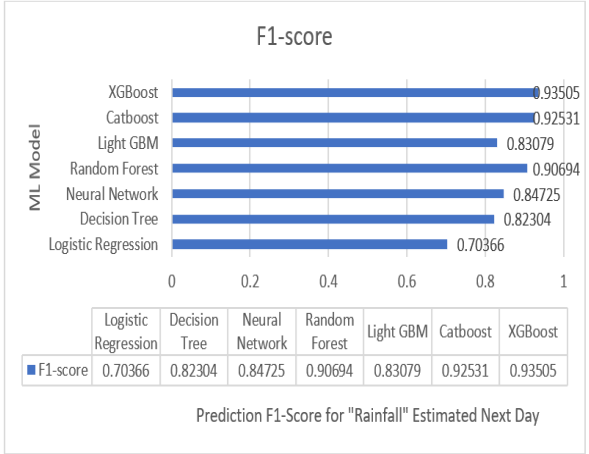



Fig. 5 - Prediction F1-Score for different ML Models




## 5. Conclusion and Future Scope

Based on the above analysis we could note down following observation that XGboost ML model performs better than the other models considered in the study. The prediction of rainfall in the present proposal species whether there will be rainfall or not the next day. For the purpose of prediction, we made use of secondary data which was already available. Accuracy, precision, recall and F1-score obtained was better for XGboost as compared to other models. The future scope is to enhance the prediction in the range of nowcasting to long range forecasting. We also look forward to obtain the rainfall estimated in the prediction list which will be helpful in many domains such as aviation, agriculture, etc. Our interest is to apply the results of rainfall prediction identify the crop yield and crop health so that majority of the farmers can take advantage of this work.

## References

Kaushik Datta, Gouthaman P, "Rainfall prediction using machine learning and neural network", *International Journal of Recent Technology and Engineering (IJRTE)*, Vol. 9, Issue 1, May 2020, pp. 1954-1961.

Moulana Mohammed, Roshitha Kolapalli, Niharika Golla, Siva Maturi, "Prediction of Rainfall Using Machine Learning Techniques", *International Journal of Scientific and Technology Research*, Vol. 9, Issue 01, Jan 2020, pp. 3236-3240.

Bala Sai Tarun G., Sriram J.V., Teja Sreenivas K., Santhi M.V.B.T, "Rainfall Prediction using Machine Learning Techniques", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Vol. 8, Issue 7, May 2019, 957-963.

Gujanatti Rudrappa, Nataraj Vijapur, Sushant Jadhav, Prabhakar Manage, "Cloud Classification Using Ground Based Images Using CBIR and K-Means Clustering", *Biosci. Biotechnol. Res. Commun.* Vol. 13, Issue 13, Dec 2020, 95–99.

G. Rudrappa and N. Vijapur, "Cloud Classification using K-Means Clustering and Content based Image Retrieval Technique," 2020 *International Conference on Communication and Signal Processing (ICCSP),* 2020, 0700-0704, doi: 10.1109/ICCSP48568.2020.9182211.

Girish L., Gangadhar S., Bharath T.R., Balaji K.T, "Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning", *National Conference on Technology for Rural Development (NCTFRD),* 2018, 61-65.

Aakash Parmar, Kinjal Mistree, Mithila Sompura, "Machine Learning Techniques for Rainfall Prediction:", *International Conference on Innovations in Information Embedded and Communication Systems (ICIIECS),* March 2017.

Prabakaran S., Naveen Kumar P., Sai Mani Tarun P, "Rainfall prediction using modified linear regression", *ARPN Journal of Engineering and Applied Sciences,* 12(12), 2017, 3715-3718.

George Box E.P., Gwilym Jenkins M., Gregory Reinsel C., Greta Ljung M, *"Time Series Analysis: Forecasting and Control",* 5th Edition. John Wiley & Sons. 2015.

Etuk, E.H., Mohamed T.M, "Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by sarima Methods", *International Journal of Scientific Research in Knowledge.* 2(7), 2014, 320-327.

https://rdrr.io/cran/rattle.data/man/weatherAUS.html#heading-1 (Accessed on 03/04/2021)