# Performance Analysis and Assessment of Machine Learning Algorithms for Prediction of Rainfall

**A. Gayathri , B. Uma Devi , Dr. T. Ramaswamy**

*B.tech 4th year, ECE* SNIST Hyderabad, India

*B.tech 4th year, ECE* SNIST Hyderabad, India

*Associate professor, ECE* SNIST Hyderabad, India

gayathriatmakur@gmail.com , bhukyaumadevi2018@gmail.com, ramaswamyt@sreenidhi.edu.in

**Abstract—** Agriculture plays a significant role in the Indian economy. Rainfall is crucial to agriculture, but forecasting it has become a major challenge in recent years. The ability to anticipate rainfall accurately enables farmers to better plan for their crops and take preventative measures. Global warming not only affects people and the environment negatively, but it also hastens climate change. Its climate is warming, and the ocean's level is increasing, which results in floods and the desertification of farmed lands. Unfavourable climate change causes premature and senseless amounts of precipitation. To grasp the necessity of changing the environment and its parameters, such as temperature, moistness, precipitation, and wind speed, which ultimately corresponds to precipitation projection. Rainfall is difficult to forecast because it also depends on your location

**Keywords**—*Rainfall forecasting, machine learning algorithms, R^2 score, MAE, MSE, RMSE.*

## I. Introduction

Rainfall projection plays a crucial role in human life and is necessary worldwide. The responsibility of analyzing the frequency of rainfall often falls on meteorological departments. However, accurately predicting rainfall has become increasingly challenging due to the changing conditions of the atmosphere. While it is possible to make educated guesses about the amount of rain that may fall during the summer and rainy seasons, it is essential to explore adaptable rainfall prediction algorithms. This need has led to the emergence of gifted and viable innovations, such as artificial intelligence (AI).

AI is an approach to controlling and extracting implicit, previously unknown, and potentially valuable information from data. It is a vast and profound field that continues to expand and be implemented in various domains. In the context of rainfall prediction, AI offers promising solutions by harnessing the power of advanced algorithms and machine learning techniques.

The application of AI in rainfall prediction involves analyzing large amounts of historical weather data, including precipitation patterns, temperature variations, wind speeds, and atmospheric pressure, among other variables. By processing and learning from this data, AI algorithms can identify intricate patterns and relationships that humans might overlook.

## II. Literature Survey

A variety of computer models and observational data are used in the modelling of rainfall forecasts. These models take into account a range of factors, such as atmospheric conditions, historical weather patterns, and other relevant variables, to generate predictions about future rainfall. By analyzing and interpreting these models and observations, meteorologists and scientists can make informed forecasts about when and where rainfall is likely to occur.

Regression is a frequent technique in business and climate forecasting, both statistically and empirically. In the context of weather prediction, regression techniques are employed to establish relationships between different meteorological variables and rainfall. These techniques analyze historical data to identify patterns and correlations, enabling scientists to develop models that can estimate future rainfall based on the observed patterns.

By employing regression methods, scientists can derive mathematical equations that capture the relationships between meteorological variables and rainfall. These equations can then be utilized to make predictions about future rainfall events. Regression analysis plays a crucial role in weather forecasting as it allows scientists to quantify rainfall and improve accuracy predictions.

El-shafie et al. used a number of statistical measures, such as correlation coefficients and These measurements make it possible to assess the dependability . By comparing the statistical indexes obtained from their proposed model with those from alternative approaches, the researchers were able to determine the effectiveness of their ANN-based rainfall-runoff forecasting model.

## III. RESEARCH METHODOLOGY

### *A. Data Acquisition*

The India Meteorological Department (IMD), a government agency in charge of meteorological observations in India, provided the raw data utilized for analysis and modelling at the beginning of the study. The IMD had recorded a variety of meteorological measurements and observations in this raw data.

By leveraging the collected raw data from the IMD, the study aimed to derive meaningful insights and develop models that could contribute to improved forecasting and understanding of meteorological phenomena, ultimately benefiting various sectors that rely on accurate weather information.

### B. Data Pre-Processing

Preparing the raw data and making it appropriate for an AI model is known as information preparation. This is the first and most crucial stage in developing a machine learning mWe must first construct a dataset since a machine learning model is fully dependent on data. The prepared collection of data for a particular issue is the dataset.

For instance, the dataset needed to create a machine learning model for commercial use we need for a liver patient will be different from the one we need for a machine learning model.

Exploratory Data Analysis (EDA):

Prior to applying machine learning methods, exploratory data analysis (EDA) is conducted to gain a better understanding of the dataset.

EDA involves utilizing various statistical and visualization techniques to explore and analyze the data. Its purpose is to uncover patterns, relationships, and key insights within the dataset.

Data Cleaning:

Data cleaning is a subset of data pre-processing. Some of the processes that fall under the category of "data cleaning" were employed to eliminate redundant data, outliers, and missing information.

Feature Engineering:

In the pre-processing phase of machine learning, there is a crucial step known as feature engineering. This process involves transforming raw data into a set of relevant features that can be utilized to build a predictive model using statistical modeling or machine learning algorithms.

The primary objective of feature engineering is to enhance the performance and effectiveness of the models.

### C. Build the machine learning model

Training algorithms employ either labelled or unlabelled data, or a combination of the two, to build machine learning models. A machine learning algorithm can be trained and produced in three primary ways:
Supervised instruction: Regulated learning happens when a calculation is prepared utilizing "named information", or information that is labelled with a mark so a calculation can effectively gain from it. The eventual machine learning model learns how to classify data in the manner desired by the researcher by training an algorithm with labelled data.

Learning without supervision: An algorithm is trained using unlabelled data in unsupervised learning. The algorithm finds patterns in the data and creates its own data clusters during this procedure. Unaided learning is useful for specialists who are hoping to find designs in information that are presently obscure to them.

### D. Predict the target variable

When evaluating a trained model's performance, it is common to use a test dataset that lacks the target variable. The target variable is the feature of the dataset that, using the remaining features, the model hopes to anticipate or predict. It represents the variable of interest for the client or the problem being addressed.

## IV. Machine Learning models

The field of predictive modelling, often known as machine learning, is primarily focused with reducing model error. Linear, Ensemble, and trees are the three categories into which all machine learning models for regression may be placed.

Linear models are Machine Learning (ML) models with their own fundamental machine learning structure, such as Lasso, Ridge, and Elastic Net Regressors, Decision Tree Regressor is a part of the tree models

### A. Ridge Regression

Ridge regression may frequently lower the variability of the estimated coefficients and provide a more stable and understandable model by tolerating a little degree of estimation bias. To assist in determining the value of the ridge parameter, a ridge trace and plot of the variance inflation factors (VIF) are supplied.

### B. Lasso Regression

When there are more unknown parameters to be estimated (p) than there are observations (n), this situation is referred described as being "high- dimensional" (p n). Traditional statistical approaches cannot handle situations where p n because they presuppose numerous observations and a small number of unknown variables.

Ensemble Methods

Except for noise, which is a basic mistake, ensemble models assist in reducing these issues.An ensemble model is one that includes models such as bagging and boosting.

### C. Gradient Boosting

A robust machine-learning approach called gradient boosting has been used in a wide variety of real-world applications. Boosting technique to make poor learners dynamic learners. For issues involving regression and classification, use the gradient boosting technique.

A robust machine-learning approach called gradient boosting has been used in a wide variety of real-world applications

### D. Random Forest Regression

Arbitary forestland is the name given to the arrangement of chosen trees, each of which is done uniquely. The GINI indicator is used by the decision tree to partition characteristics.

## V. RESULTS

### A. Performance Measure

In this part, we investigate the regression of machine literacy methods. We describe the lariat, crest, and elastic net of direct models, the arbitrary timber regressor, the grade descent regressor of ensemble models, and the decision trees of trees before comparing the outcomes. As stated, there are 4188 total conditions in the exam, of which 80 are chosen for preparation (out of 3350 total information tests) and 20 are chosen for testing (out of 838 total information tests). The questions in this paper were drawn from one of the 838 information exams. Regression techniques are evaluated using the R2 score, Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE).

R Squared ($R^2$ or R^2 score):

R2 expresses "how well a regression line predicts factual values" R-squared measures the percentage of the target variable difference that the direct model captures. The R-squared value is between 0 and 100. If the R-squared number is substantial, which indicates that it is near to 100 or even "0," the model correctly fits the data. However, if the R-squared value is unacceptably low, the model does not accurately match the data and provides inaccurate predictions. Y is therefore the average of the true values and the values from the best-fit line.

Mean Absolute Error (MAE):

There are several methods for estimating the model's performance. MAE is one of the standards used to define and rate the quality of a machine literacy model. Where yi is the actual value and i is the prognosticated value, the error is defined in MAE as a normal of the outright divergence between the actual rates and the anticipated rates.

Mean Squared Error (MSE):

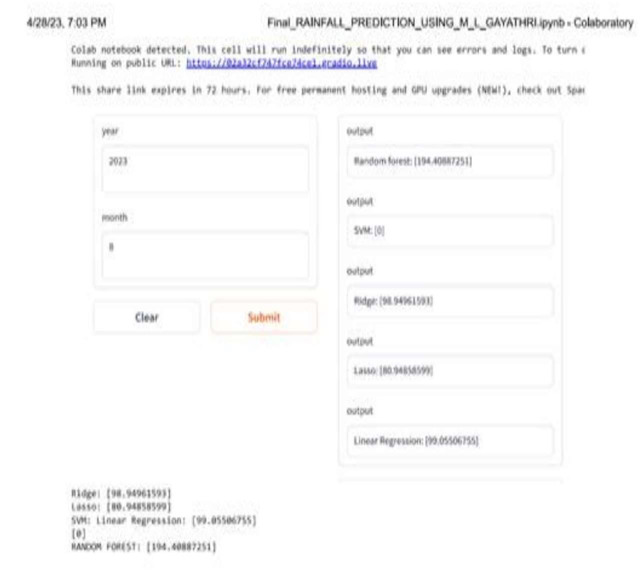
The most common statistic used to assess regression issues is "Mean Squared Error". The average squared error between the actual and predicted values for each point is calculated in detail. However, if the average squared error values are high, the model is also poorer.

Root Mean Squared Error (RMSE):

The square root of MSE is RMSE. To make the size of the offences and the size of the target equal, the square root formula is used.

### B. Analysis of Results

This section looks at the results of an extensive investigation on many colourful AI(ML) computations, including Rope, Edge, Flexible Net of Straight Models, Arbitrary Woodland, Inclination Helping Regressor of Outfit Models, and Choice Tree Regressor for rush datasets. The performance metrics of ML outcomes on downfall datasets are shown. According to the lasso regression model performs better in terms of R2 Score.



4/28/23, 7:03 PM Final_RAINFALL_PREDICTION_USING_M_L_GAYATHRI.ipynb - Colaboratory

```
Colab notebook detected. This cell will run indefinitely so that you can see errors and logs. To turn c
Running on public URL: https://02a32cf747fce74ce1.gradio.live

This share link expires in 72 hours. For free permanent hosting and GPU upgrades (NEW!), check out Spa
```

year: 2023

month: 8

output: Random forest: [194.40887251]

output: SVM: [0]

output: Ridge: [98.94961593]

output: Lasso: [80.94858599]

output: Linear Regression: [99.05506755]

```
Ridge: [98.94961593]
Lasso: [80.94858599]
SVM: Linear Regression: [99.05506755]
[0]
RANDOM FOREST: [194.40887251]
```

Figure 1 output of predicted rainfall using various ML algorithms

## VI. CONCLUSION

Numerous ML methods have been used to enforce automated regression for rainfall. This research presents a description and an example of how each of the six machine learning algorithms performs while regressing against the common rainfall dataset. The lasso regression approach has the lowest R2 score of the six algorithms at 80–20% of the training and confirmation dataset, with 99.21%. Away from this, all ML calculations are evaluated and altered according to the actual objective rates and expected values. Regression methods can be used to improve the point's delicateness.

## REFERENCES

[1] N. Gnana Sankaran, E. Ramaraj, “A Multiple Linear Regression Model to Predict Rainfall Using Indian Meteorological Data”, International Journal of Advanced Science and Technology (IJAST) Vol. 29, No. 8s, (2020), pp. 746-758.

[2] Irasema Alcntara-Ayala. Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. Geomorphology, 47(24):107–124, October 2002.

[3] Neville Nicholls. Atmospheric and Climatic Hazards: Improved Monitoring and Prediction for Disaster Mitigation. Natural Hazards, 23(2-3):137–155, March 2001

[4] Puneet Sharma and Nadim Chishty, "Machine Learning-Based Modelling of Human Panther Interactions in Aravalli Hills of Southern Rajasthan”, Indian Journal of Ecology 46(1): 126-131.

[5] A.El-shafie, M.Mukhlisin, Ali A. Najah and M.R. Taha, "Performance of artificial neural network and regression techniques

for rainfall-runoff prediction", International Journal of the Physical Science vol 6(8), 18 April 2011.