# Prediction of Rainfall using Machine Learning algorithms

P.Ramkumar

*Department of Artificial Intelligence and Machine Learning*

*Sri Sairam College of Engineering*

*Anekal, Bangalore, Karnataka-562106*

*Email:ramkumarkohila@gmail.com*

Reji Thomas

*Dept of Computer Science and Engineering*

*Sri Sairam College of Engineering*

*Anekal, Bangalore, Karnataka-562106*

*Email:rejitoms@gmail.com*

**Abstract**:

*One of the pillars of the Indian economy is the agricultural sector. Even while rainfall is crucial for farmers, predicting it has become a huge challenge in recent years. If farmers can accurately estimate when it will rain, they may better plan their crops and avoid problems. Alterations to the weather are being hastened by global warming, which is having devastating effects on both humans and the natural world. Because of the warming air and rising sea levels, floods are becoming more common and drought is becoming more common in farmed fields. Bad climate change causes excessive rainfall that is neither seasonal nor appropriate. The ability to forecast precipitation is a powerful tool for understanding weather patterns. The overarching goal of this research is to help clients in the agricultural, research, and power generation sectors, among others, understand the importance of climate change and the parameters that influence it, such as temperature, humidity, precipitation, wind speed, and rainfall projections. It is difficult to forecast rainfall because it is also dependent on geographic regions. Machine learning, a dynamic branch of artificial intelligence, aids in weather prediction. In order to forecast the weather, this study will use a dataset with several attributes from the UCI repository. Building a more accurate method for predicting rainfall using Machine Learning classification algorithms is the primary goal of this research.*

**Keywords:-** *Machine Learning, Classification algorithms Rainfall Prediction system*

## 1.**Introduction**

One of the most important aspects of human existence is the ability to predict when and how much rain will fall. Analysing the regularity of rainfall with caution is a heavy burden for the meteorological department. Precise rainfall forecasting is challenging due to the volatility of atmospheric conditions. An attempt at summery and rainy season rainfall prediction is speculative at best. Consequently, it is essential to investigate the algorithms that can be adjusted to forecast rainfall. "Machine Intelligence is a way of processing and extraction of implicit, previously unsuspected

and recognised and potentially useful information about data." This is a description of one capable and efficient technology. The size and depth of the discipline of machine learning, as well as its potential applications, are constantly expanding.

In order to generate predictions and determine the accuracy of a dataset, machine learning makes use of a variety of classifiers, including supervised, unsupervised, and ensemble learning. That information will be useful for our Rainfall Prediction System project, which we are working on. To identify the best model, we examine many Machine Learning techniques, including Decision Tree, Random Forest, K-Nearest Neighbour, and Logistic Regression. The rainfall information stored at the University of California, Irvine, is utilised in this case. All of the current methods of classification are reviewed and compared in this comprehensive study. Considerations for future research and potential avenues for progress are also discussed in the report.

Predicting a location's rainfall using user-supplied criteria is the major goal of this research post. Day, place, temperature range, humidity, wind speed and direction, evaporation rate, and other variables are all part of the parameters. Logistics regression, KNN, Decision Tree, and Random Forest are the four techniques that are used to train these rainfall attributes. The two most effective algorithms, Random Forest and KNN, provide an accuracy of about 88%. Finally, we will make a prediction about the state of the rainfall in that area.

## 2.Literature Survey

Considering the many methods proposed by authors, this research aims to build a real-time rainfall forecast system that improves upon and eliminates the drawbacks of existing methods. Rainfall in India's Udupi district, located in the state of Karnataka, is predicted using the method [1]. The method of BPNN with cascade feedforward neural networks is employed. Comparing the network to BPNN, it demonstrates superior accuracy. For long-term rainfall forecasting, this model might not be very reliable.

Its structure [2] G. Geetha and R. Selvaraj factored in a number of meteorological variables, including high and low temperatures, relative humidity, wind speed and direction, and monthly rainfall forecasts for the Chennai area, using an ANN model. After looking over the numbers, they were able to forecast how much rain will fall each week in several areas of Chennai. Compared to numerous linear regression models, ANN has better accuracy when used for prediction. Both the forward pass and the backward pass are utilised by this method. By use of the network, data is transmitted from the input layer to the subsequent layer via the forward layer. The last step is to analyse the results from the previous layer and then produce the outcome at the backword layer. In their suggested paper, the authors [3] presented a system for predicting rainfall using the KNN deep learning technique. The total number of nearest neighbours is utilized to ascertain the class label for unknown data, and a single K-value is provided for this purpose. By using KNN, we can identify the category or class of a given dataset by grouping parameters with similar characteristics into the same sort of cluster. Neither classification nor regression training of this technique is time-consuming. The accuracy of this method could be compromised if the wrong value of K is chosen.

## 3. Proposed Methodology

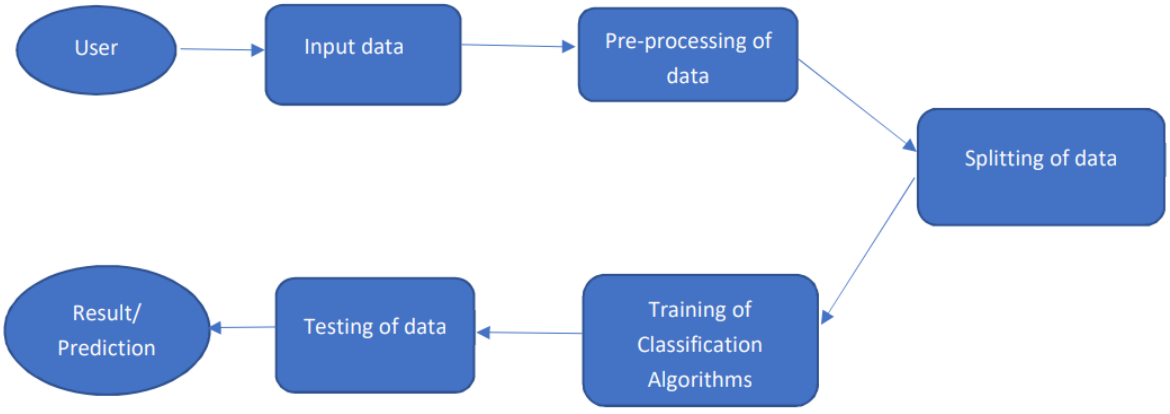


Figure3.1.System Architecture

The figure3.1 depicts that the flow diagram of prediction of rain flow using machine learning techniques. In this system the input data is pre processed and then split it into two such as training and testing data. Based on the machine learning methods the rain flow will be predicted.

3.1. Analyzing and Exploring Data:

If you want your predictions to be accurate and reliable, you need to undertake data analysis to ensure that your results are close to the mark. You can only be sure that the data was collected accurately after verifying and checking the raw data for anomalies. Data that contains features that aren't essential to the prediction model can also be located with its help.

3.2.Pre-Processing of data

As a data mining approach, data pre-processing transforms unstructured and messy data into a more comprehensible and applicable format for the model. In addition to having numerous mistakes and missing features, raw data is unreliable and incomplete. Information gleaned from our data exploration and analysis has shown us that our model's raw data has numerous null values that need to be replaced with their respective means. Furthermore, we can deal with the missing values by removing any unnecessary columns or rows. Because models rely on computations and mathematical equations, it is essential to encode categorical data in order to transform it into numerical form. Another aspect of pre-processing is feature selection, wherein we pick out the features that will actually help our rainfall prediction model, which in turn shortens the training period and improves the model's accuracy. At last in the pre-processing pipeline, feature scaling ensures that all independent variables are within their respective ranges and that no one variable has a significant impact on the others.

## 4.Methods

Cleaning, preprocessing, and organising acquired meteorological data are the initial steps in the suggested model. The Indian Meteorological Department has established a system for classifying rainfall data. Our method for predicting future rainfall using ML classification algorithms is detailed in this publication. The pre-processed data is divided into two parts: 30% for testing and 70% for training. We apply four separate machine learning algorithms to the partitioned data, evaluate each output, and then show you the correct outcome. The next section details how each classifier functions.

4.1.Logistic Regression
One supervised learning classification approach that can be utilized to forecast the likelihood of a target variable given information is logistic regression. With a binary nature like that of a target or dependent variable, there can be only two possible outcomes: success or failure.

4.2. K-Nearest Neighbor (K-NN):
As an example of a basic supervised learning-based machine learning algorithm, K-Nearest Neighbour is among the most basic. In order to classify new cases, the K-NN algorithm looks at how similar they are to existing examples and assigns them to the category that is most closely connected to those categories. Based on the objects' nearest neighbours, it sorts them into categories. After you name a point, it will group similar points together and utilise them to mark another point. Clustering comparable data makes it possible to fill in missing values with K-NN. The data set is subjected to ML algorithms once these missing values are filled. You can improve the accuracy by using different combinations of these methods.

4.3. Random Forest
A robust ensemble learning method, the Random Forest algorithm trains by building a large number of decision trees. One random feature subset and one bootstrapped dataset sample are used to train each tree. For regression tasks, the final output is produced by averaging the predictions from individual trees; for classification tasks, it is produced via voting. In order to improve the model's predicted accuracy and resilience, this strategy reduces variance and helps with overfitting. The significance of features can be better understood with the use of Random Forest, which in turn helps to find strong predictors of the target variable. When compared to individual decision trees, it is less affected by noise and outliers, has a low processing footprint, and can handle big datasets with ease. Because of its adaptability and efficacy, Random Forest is a favorite for predictive modeling jobs and has found numerous uses in fields as diverse as healthcare, finance, and the environment.

4.4. Decision tree
Machine learning's Decision Tree approach builds a decision-node tree structure by recursively partitioning the dataset according to feature values; it's both flexible and interpretable. Optimal splits are achieved at each node by picking a feature, like Gini impurity or entropy, that maximises information gain or minimises impurity. This procedure keeps running until some endpoint is reached, like a certain depth or a certain minimum number of samples per leaf. Because

of their visual nature and ease of understanding, decision trees are useful for regression and classification tasks alike. Their inability to generalise to new data and propensity for overfitting are issues, particularly when dealing with complicated datasets. It is common practice to use ensemble approaches such as Random Forests and Gradient Boosting Machines (GBM) to solve these problems with decision trees while keeping their interpretability. In spite of their flaws, decision trees are nevertheless widely used because of their transparency, simplicity, and capacity to represent intricate decision boundaries.

4.5.Accuracy

A measured or anticipated value is considered accurate if it approaches the true or known value within a reasonable margin of error. A model's or a measurement's accuracy in representing reality is quantified by this metric. Correctly predicted instances (including true positives and true negatives) as a percentage of total instances is called accuracy in classification tasks. A 90% accuracy rate would be achieved, for instance, if the model accurately detects 90% of the cases. When one class is significantly more numerous than another in a dataset, accuracy may not be indicative of true performance.

4.6.Precision

Precision is a measure of how well a model can make positive predictions. It is sometimes called positive predictive value. The ratio of the model's real positive predictions to its total number of positive predictions is what this metric measures. For binary classification tasks, where accuracy measures how well a model can avoid making incorrect predictions, precision is crucial. In terms of the ratio of false positives to total predictions, a high precision shows that the model is quite accurate.

## 5. Results and Discussions

Table 5.1Accuarcy of data

| Techniques | Accuracy of classification | Value of Precision |
|---|---|---|
| Logistic Regression | 80.36 | 0.742 |
| KNN | 83.25 | 0.781 |
| Decision Tree | 60.76 | 0.26 |
| Random Forest | 88.26 | 0.864 |

The table 5.1 depicts that the performance of prediction of rain flow using machine learning algorithms. It is identified that compare to all the machine learning methods, random forest provides a high accuracy and precision values.

## 6. Conclusion

Various ML approaches that can be used for rainfall prediction are going to be defined. Using a smaller set of features and tests to create a more efficient and accurate model is the focus of this study. It all starts with pre-processing the data before it's incorporated in the model. Classification algorithms with the highest efficiency are Random Forest classifier (about 88%) and K-Nearest Neighbour (87%). But with just 73% accuracy, the Decision Tree classifier comes out on the bottom. Additional machine learning methods, including ensemble methods, time series, clustering, and association rules, can be explored in this study. More complicated and combined models are required to achieve better accuracy for rainfall forecast systems, bearing in mind the study's limitations. Research can also be designed with more precise and accurate calculation rates in mind by utilising more articulate monitoring for certain areas and developing such a model for massive datasets.

## REFERENCES

1. Kumar Abhishek. Abhay Kumar, Rajeev Ranjan, Sarthak Kumar," A Rainfall Prediction Model using Artificial Neural Network", 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC2012), pp. 82-87, 2012.
2. G. Geetha and R. S. Selvaraj, “Prediction of monthly rainfall in Chennai using Back Propagation Neural Network model,” Int. J. of Eng. Sci. and Technology, vol. 3, no. 1, pp. 211 213, 2011.
3. Zahoor Jan, Muhammad Abrar, Shariq Bashir and Anwar M Mirza, "Seasonal to interannual climate prediction using data mining KNN technique", International Multi-Topic Conference, pp. 40-51, 2008.

4. Elia Georgiana Petre, "A decision tree for weather prediction", Seria Matematica - Informatica] – Fizic, no. 1, pp. 77-82, 2009.
5. Gupta D, Ghose U. A Comparative Study of Classification Algorithms for Forecasting Rainfall. IEEE. 2015.
6. Wang J, Su X. An improved K-Means clustering algorithm. IEEE. 2014.
7. Rajeevan, M., Pai, D. S., Anil Kumar, R. & Lal, B. New statistical models for long-range forecasting of southwest monsoon rainfall over India. Clim. Dyn. 28, 813–828 (2007).
8. Mishra, V., Smoliak, B. V., Lettenmaier, D. P. & Wallace, J. M. A prominent pattern of year-to-year variability in Indian Summer Monsoon Rainfall. Proc. Natl Acad. Sci. USA 109, 7213–7217 (2012).
9. Thirumalai, C., Harsha, K. S., Deepak, M. L., & Krishna, K. C. (2017). Heuristic prediction of rainfall using machine learning techniques. 2017 International Conference on Trends in Electronics and Informatics (ICEI).