iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42876

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Rainfall Prediction using Machine Learning

Vikas Kumar[1], Vishal Kumar Yadav[2], Er. Sandeep Dubey[3]
[1, 2]*Department of Computer Science and Engineering*
[3]*Guide*

Abstract: ***In India, Agriculture contributes major role to Indian economy. For agriculture, Rainfall is important but during these days' rainfall prediction has become a major challenging problem. Good prediction of rainfall provides knowledge and know in advance to take precautions and have better strategy about theirs crops. Global warming is also having severe effect on nature as well as mankind and it accelerates the change in climatic conditions. Because of its air is getting warmer and level of ocean is rising, leads to flood and cultivated field is changing into drought. Due to adverse climatic change leads to unseasonable and unreasonable amount of rainfall. To predict Rainfall is one of the best techniques to know about rainfall and climate. The main aim of this study revolves around providing correct climate description to the clients from various perspectives like agriculture, researchers, generation of power etc. to grasp the need of transformation in climate and its parameters like temperature, humidity, precipitation, wind speed that eventually directs to projection of rainfall. Rainfall also depends on geographic locations hence is an arduous task to predict. Machine Learning is the evolving subset of an AI, that helps in predicting the rainfall. In this research paper, we will be using UCI repository dataset with multiple attributes for predicting the rainfall. The main aim of this study is to develop the rainfall prediction system and predict the rainfall with better accuracy with the use of Machine Learning classification algorithms.***

Keywords: Rainfall Prediction system, Machine Learning, Dataset, Classification algorithms.

I. INTRODUCTION

Rainfall projection is utmost necessary all over world and it plays a key role in human life. It's cumbersome responsibility of meteorological department to analyze the frequency of rainfall with precariousness. It is difficult to forecast the rainfall precisely with varying atmospheric condition. It is conjectured to predict the rainfall for both summer and rainy seasons. This is the primary reason because of this there is necessity to analyse about the algorithms adaptable for rainfall prediction. One of such skilled and effective technologies is Machine Learning, "Machine Learning is a way of manipulating and extraction of implicit, previously unknown and known and potential useful information about data". Machine Learning is colossal and deep field and its scope and implementation is increasing day by day.

Machine learning covers various classifiers of Supervised, Unsupervised and Ensemble Learning which are used to predict and find the accuracy of the given dataset. We can use that knowledge in our project of Rainfall Prediction System as it will help a lot of people. Various Machine Learning algorithms such as Logistic Regression, Decision Tree, K-Nearest Neighbor, Random Forest are compared to find the most accurate model. Here the rainfall dataset from the UCI repository is used. In this research a discussion and comparison of the existing classification techniques is made. The paper also mentions scope of future research and different advancement possibilities.

The objective of this research paper is to predict the Rainfall of a location based on input parameters that will be provided by the user. The parameters include date, location, maximum temperature, minimum temperature, humidity, wind direction, evaporation etc. These rainfall attributes are trained under four algorithms: Logistic regression, KNN, Decision Tree and Random Forest. Most efficient of these algorithms are Random Forest and KNN which give us the accuracy of approximately 88%. And, finally we will predict the rainfall status of that particular place.

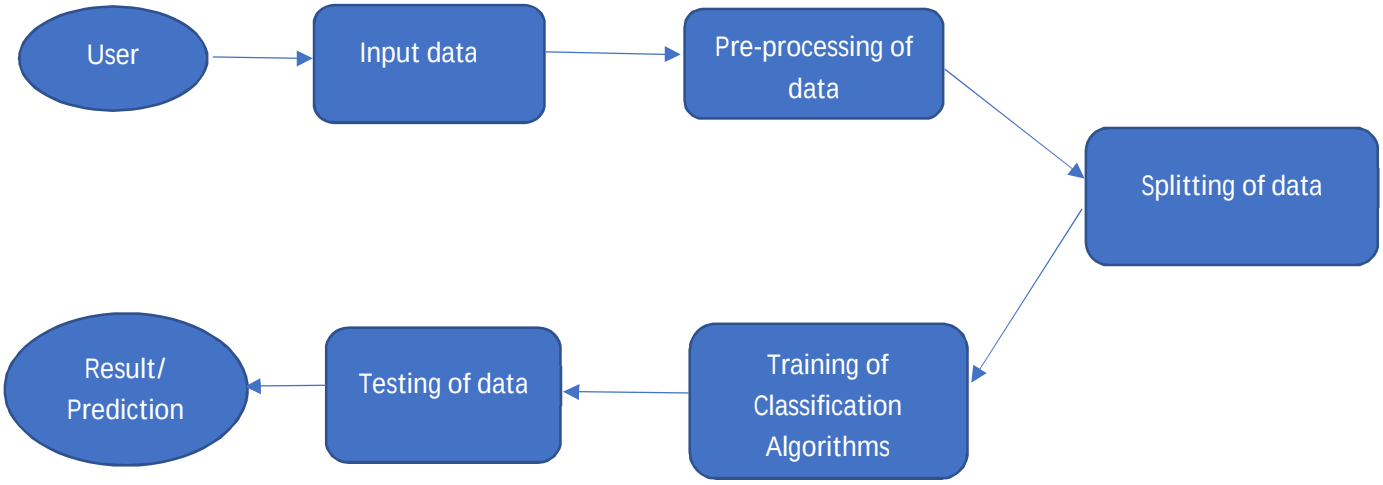
II. LITERATURE REVIEW

The primary aim of this paper is to study the different approaches as given by authors and to develop a real time rainfall prediction system which overcomes the shortcomings of previous methods and to give the best and accurate solution. The system [1] predicts the rainfall of Udupi district from the Karnataka state of India. BPNN with cascade feed forward neural networks technique is used. The network shows better accuracy when compared to BPNN. This system might not work accurately for a long period prediction of rainfall.

©IJRASET: All Rights are Reserved

The system [2] G. Geetha and R. Selvaraj used ANN model for predicting monthly rainfall over Chennai region and took various attributes of weather such as maximum and minimum temperature, and relative humidity, wind speed, wind direction. They analysed the data and predicted weekly rainfall over selected regions of Chennai. Prediction using ANN gives good accuracy than multiple linear regression model. This algorithm works on two passes: forward pass and backward pass. Input is passed to the forward layer and it is propagated to next layer through network. Finally, outcome is produced at backword layer after analysing the result of previous layer. Paper proposed by [3] introduced rainfall prediction system using deep mining KNN technique. A single K value is given which is used to find the total number of nearest neighbors that helps to determine the class label for unknown data. Similar parameters are clustered into same type of cluster and thus with the help of KNN we determine the class or category of a specific datasets. This algorithm does not require time for training of classification or regression. This system may not lead to good accuracy if the incorrect value of K is picked.

III. PROPOSED METHODOLOGY



A. Data Exploration and Analysis

Data analysis is done to achieve certainty of future result to be close so that prediction is valid and correctly interpreted. This certainty can be gained only after raw data is verified and checked for abnormality thus ensuring that the data was gathered without any errors. It also helps in finding the data which contains irrelevant features for prediction model.

B. Data Pre-processing

Data pre-processing is a data mining technique that converts raw and inconsistent data into useful understandable format for the model. Raw data is inconsistent and incomplete and contains missing features along with many errors. As per data exploration and analysis we have learned that raw data for our model contains many null values which must be replaced with their mean value. We can also handle the missing values either by deleting irrelevant column or row. Encoding of categorical data is done as model is based on mathematical equations and calculations hence it is necessary to convert these categorical data into numeric. Feature selection is also the part of pre-processing in which we select only those features which contributes to our rainfall prediction model thus helps in reducing training time and increases accuracy of the model. Feature scaling is the final stage in pre-processing in independent variables are brought into specific range so that no any variable dominates the other variable.

C. Modelling

Initially in the proposed model, redeemed weather data is cleaned, then it is pre-processed and then arranged. Finally, rainfall data is designated into various categories as per Indian Meteorological Department guidelines. In this paper we have come up with an approach for the prediction of rainfall using Machine Learning classification algorithms. The pre-processed data is segregated into 70% training and 30% for testing. Four different Machine Learning Algorithms are applied on the portioned data and after that each result is analysed and final accurate result is displayed. The working of the individual classifiers is explained in the proceeding section.

1) *Logistic Regression:* Logistic regression is a supervised learning classification algorithm used to predict the probability of a given target variable. The nature of target or dependent variable is branched, there would be only two possibility of classes 0 for failure and 1 for success.
2) *K-Nearest Neighbor (K-NN):* K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm considers the similarity between the new case/data and available cases and put the new case into the category that is mostly related to the available categories. It classifies objects dependent on nearest neighbor. It groups the named points and uses them on how to mark another point. The similar data is clustered and it is possible to fill the null values of data using K-NN. As soon as these missing values are filled, we apply ML techniques to the data set. It's possible to obtain better accuracy by utilizing various combinations of these algorithms.
3) *Random Forest:* Random Forest is a supervised learning algorithm which is used for both classification as well as regression thus by creating decision trees on the data samples.

1) Step 1-There is a selection of random samples from a given dataset.
2) Step 2 – It constructs decision tree for each data sample and then it will predict from every decision tree.
3) Step 3 – After that voting will be performed on each predicted result.
4) Step 4 − At last, select the most voted prediction result as the final prediction result.

5) *Decision Tree:* This classification algorithm that works on categorical as well as numerical data is a Decision tree algorithm. It creates tree-like structures and is very easy to implement, analyse the data in tree-shaped graph.

This algorithm helps in splitting the data into two or more related sets based on the most important indicators. First, we calculate the entropy of each attribute and then the data is divided, with predictors having maximum information gain or minimum entropy: The results obtained are easier to read and interpret. This algorithm has higher accuracy in comparison to other algorithms as it analyses the dataset in the tree-like graph.

D. Evaluation

1) *Accuracy:* It is the ratio of number of correct outputs to the total number of input samples.
2) *Precision:* It is the number of correct positive correct results divided by the number of positive results predicted by the classifiers.

IV. RESULT AND ANALYSIS

The focus of this research paper is to design a model and analyse the performance of various Machine Learning algorithms and predict the most accurate algorithm for the predicting of rainfall. This research was done using techniques of Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbor on the dataset. For the experimental purpose we have given the actual real time values of maximum and minimum temperature, relative humidity, wind speed etc. Dataset was segregated into training and testing data and after those models were trained and the accuracy score was noted and analysed before final prediction. A comparison of the performance of the algorithms are represented below and their accuracy scores are shown in the table.

Method	Classification Accuracy	Precision
Random Forest	88.21	0.844
KNN (n=27)	87.36	0.791
Decision Tree	73.67	0.16
Logistic Regression	84.63	0.732

Fig.1 Accuracy on 30% test data.

V. ADVANTAGES

1) Water resources can be managed efficiently by using rainfall prediction system.
2) Regions can be evacuated if flood are expected.
3) It helps in taking appropriate measures to efficiently manage water resources, crop productivity and no wastage of any resources.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

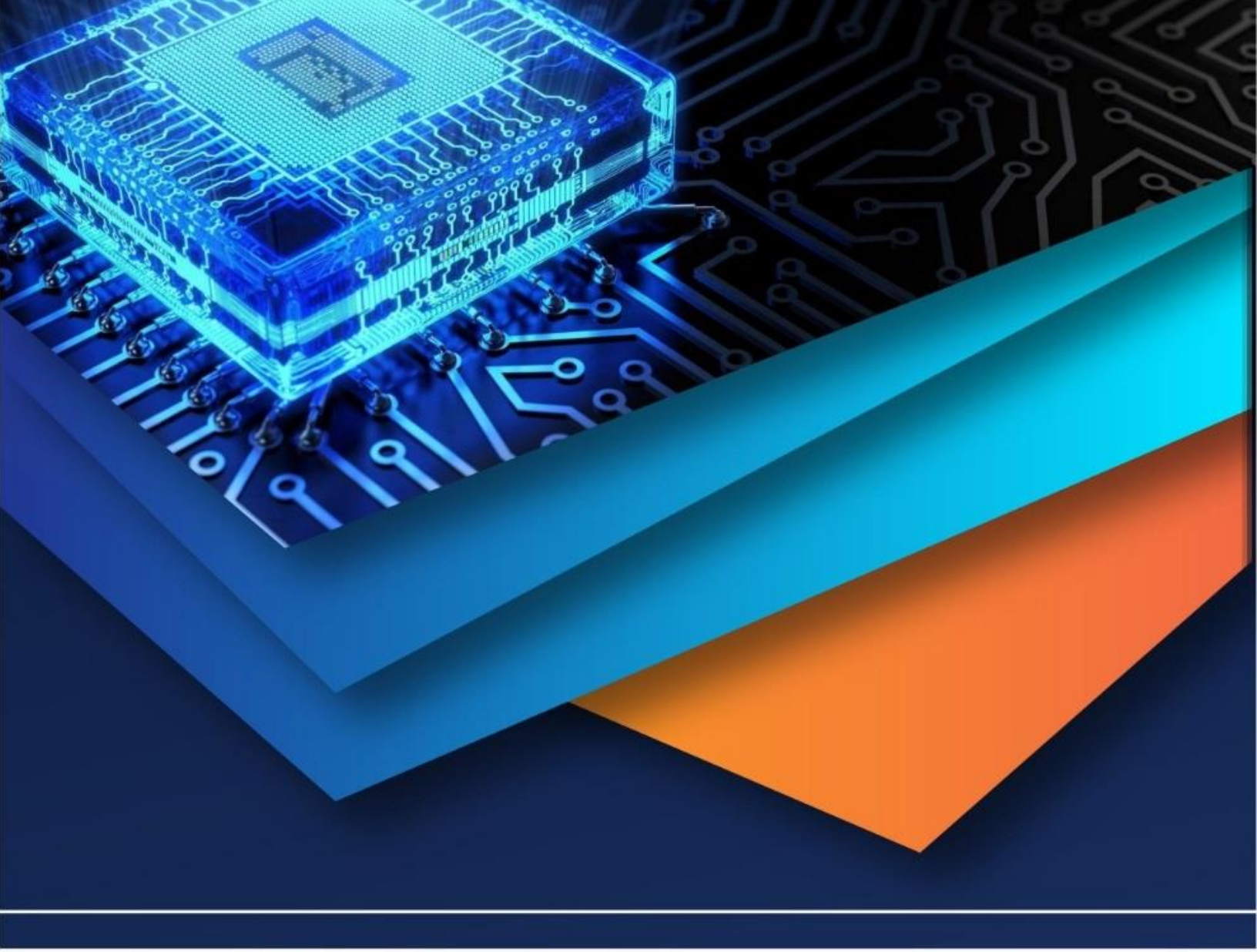
VI. CONCLUSION

The overall aim is to define various ML techniques that are useful in predicting rainfall. The goal of this research is to design accurate and efficient model by applying lesser number of attributes and tests. Firstly, the data is pre-processed and then it is used in the model. K-Nearest Neighbor with 87% and Random Forest classifier with approximately 88% are the most efficient classification algorithms. However, Decision Tree classifier gives the least accuracy with 73%. We can further expand this research covering other ML techniques such as time series, clustering and association rules and other ensemble techniques. Taking into consideration the limitations of this study, there is a need to build more complex and combination of models to get higher accuracy for rainfall prediction system. Study can also be formulated using greater articulate monitoring for particular area and create this kind of model for enormous dataset so that calculation rate can be increased with better precision and with more accuracy.

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).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)