Enhancing Prediction Accuracy of Rainfall Using Machine Learning And Forecasting Methods

1st Niharika Jain
Computer Science Department Meerut Institute of Engineering and Technology
Meerut, India
niharika.jain.cs.2019@miet.ac.in
2nd Nikita Jain
Computer Science Department Meerut Institute of Engineering and Technology
Meerut, India
nikita.jain.cs.2019@miet.ac.in
3rd Shivam Sharma
Computer Science Department Meerut Institute of Engineering and Technology
Meerut, India
shivam.sharma.cs.2019@miet.ac.in
4th Rakshit Mishra
Computer Science Department Meerut Institute of Engineering and Technology
Meerut, India
rakshit.mishra.cs.2019@miet.ac.in
5th Vaishali Deshwal
Computer Science Department Meerut Institute of Engineering and Technology
Meerut, India
vaishali.deshwal@miet.ac.in'

Abstract—Agriculture plays a major part within the Indian economy. Rain is crucial for agriculture, but of late, predicting rain has become a really tough subject. An honest rain forecast provides farmers the data to set up ahead, take precautions, and have higher agricultural ways. Both nature and humanity are being severely impacted by global warming, that conjointly hastens the amendment in weather conditions. Flooding and therefore the transformation of the cultivated field into aridity are caused by the warming of its air and therefore the increasing ocean level. Unreasonable and unseasonal rain may be a result of unfavorable environmental conditions. One of the best ways for learning regarding rainfall and climate is to anticipate rain. This research's major objective is to accurately describe to clients about precipitation from a variety of perspectives, including agriculture, research, the production of electricity, etc. In order to achieve accurate prediction, variables such as temperature, humidity, precipitation, and wind speed, are used which ultimately help with generating results in rainfall prediction. The ever-evolving portion of computer science that aids with rain prediction is named Machine Learning. For the aim of predicting the rain during this analysis study, we'll use a dataset that has several properties. The most important goal of this work is to spot the best algorithmic rule for predicting rain and to try to do so accurately, we are using Machine Learning regression algorithms.

Keywords —Random Forest, Machine Learning, Neural Network, K-NN, Decision Tree, SVR and Rainfall Prediction system.

I. INTRODUCTION

Rainfall forecasting is crucial to human survival and is essential everywhere in the world. Analysis of rainfall

frequency is a difficult task that falls under the meteorological department's arduous purview. With shifting atmospheric conditions, it is impossible to predict the rainfall with accuracy. This is the main justification for why it is necessary to investigate the algorithms that may be used to forecast rainfall. The timing and amount of precipitation have great significance and can have an impact on the entire economy of a country like India, precipitation serves as the main supply of water for the majority of the agricultural industry. The ability to predict potential future events for a system is the phenomenon of prediction. Instruments on the ground and satellite-based remote sensing techniques are used to get the current weather observations. Precipitation is crucial since India's economy is heavily dependent on horticulture. A proficient and effective technique called machine learning is able to handle and retrieve implicit, previously unknowable, known, and potentially useful information about data. It is becoming more and more common to use machine learning, which is a broad and complicated field [7]. To forecast and evaluate the precision of the supplied dataset, machine learning employs a number of categories from ensemble, supervised, unsupervised learning. Given that so many individuals will find value in that information, we can use it in our project to forecast rainfall. Various algorithms, including Neural Network, Random Forest, Decision Tree, K-Nearest Neighbor and Support Vector Machine are investigated to find which machine learning method is the best at accuracy. The many methods of regression are reviewed and compared in this paper. In addition to exploring methods that have been used to forecast rainfall, machine learning, and their shortcomings, our work is primarily concerned with acknowledging the multiple meteorological factors that affect the prediction of rainfall [8] [9].

The suggested model predicts how much rain will fall by methods for forecasting using machine learning. The variation in precipitation is primarily attributed to its association with Humidity, Temperature, Wind Speed, Pressure etc. By using user-supplied input parameters for the aforementioned features, this research study will try to estimate how much rain will fall. These variables associated with rainfall are trained using five different algorithms: Decision Tree, Support Vector Machine, Random Forest, Neural Network and K Nearest Neighbor. The most effective of these algorithms are Random Forest and Decision Tree, which gives us an accuracy of approximately 0.983 and 1.00 respectively [10] [11]. This is our contribution to the problem: analyzing the precision of various approaches for forecasting precipitation using machine learning. The text in addition contains information on future research's scope and other chances for advancement.

II. LITERATURE REVIEW

In order to construct a real-time rainfall prediction system that addresses the drawbacks of earlier systems and provides the best and most accurate solution, the primary objective of this paper is to investigate the various methodologies as presented by the authors. Time-series data were employed for forecasting by the authors in [2]. The time-series data is temporal in nature because it is derived from

data sources including scientific research, financial software, GPS, weather data, and other sources. The modeling of dynamic and non linear systems using artificial neural networks (ANN) is a well established method. For one-month and two-month rainfall forecasts, a model has been constructed [12].

This strategy is more appropriate than conventional and numerical approaches, according to Deepak Ranjan Nayak and his team's paper [3] on rainfall forecast using ANN. Additionally, they looked at a few of the more widely used artificial neural networks, including the Self-Organizing Map (SOM), Radial Basis Function Network (RBFN), Support Vector Machine (SVM), and Back Propagation Network (BPN). The paper can be more helpful for those who utilize ANN to make their forecasts, while MLR, SVM, BPN, and SOM-based algorithms are better at predicting rainfall [13] [14].

In work [4] [15], the researchers suggested a brand-new method for forecasting rainfall. It was done in two steps. The collection of features is trimmed down and the potential segments for rain predictions are taken into account using the biased forward classification approach. A different Neural Network is trained for each cluster after the dataset has first been aggregated using the k-means method. The classifier connected to several statis tical efficiency assessment factors was compared to the two step prediction approach that was suggested. The Dumdum meteorological department gathered data between 1989 and 1995 in order to conduct experiments [16] [17].

In an experiment [5] they conducted, Tanveer Hurra and Shreekanth Parashar used data mining techniques to forecast when it would rain.To determine which method(s) best fit the issue, they employed the Naive Bayes algorithm, ANN, Decision Trees, Random Forest, and k-nearest neighbor algorithm.

They have further found that the best models for predicting rainfall are those based on decision trees and random forests. Random Forest, Support Vector Machine and Decision-tree are a few categorization techniques that the authors of [6] have utilized. For each method, they calculated the precision, and the values of F-measure, recall, and accuracy. Out of all, Random Forest provided an accuracy of 0.8, which is higher [18] [19].

III. WORKING

A. General Architecture

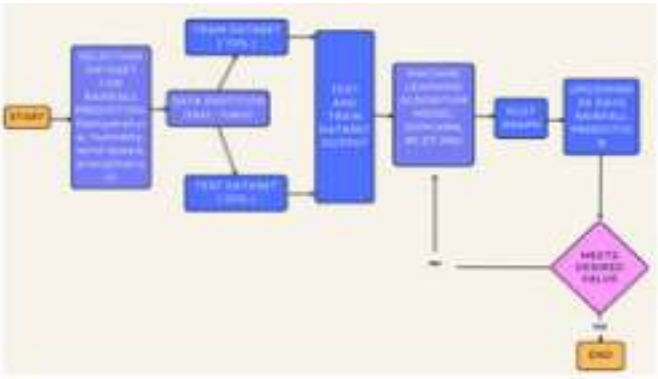

Fig. 1. Flowchart for Execution of Rainfall Predictor

B. Algorithm

1) Data Exploration and Analysis:

Data analysis is to understand and create a thorough comprehension of the information to help better understand the nature of the data.

2) Data Preprocessing:

Our model's raw data has a sizable number of null values that need to be replaced with zero, we can also take care of the missing values by deleting unnecessary columns or rows.

3) Modeling:

We offer a technique that makes use of machine learning for predicting rainfall with the help of regression techniques. 'sklearn' library is used for execution. 0.3 of the pre-processed data are used for testing, while 0.7 are used for training. The workings of each regressor is described in the section that follows.

C. Machine Learning Regression Models

1) Support Vector Regressor:

With Support Vector Regression we have the freedom to specify the level of error that is permissible in our model, and it will locate a suitable line (or hyperline in higher dimensions) that fits the data.

2) Random Forest:

Leo Breiman and Adele Cutler, the inventors of the Random Forest machine learning algorithm, combined the output of various decision trees to produce a singular outcome.

3) Decision Tree:

You can make choices based on past experiences by using a decision tree, which is a flowchart. This technique, which is more accurate than other algorithms, analyzes the information as a tree-like graph.

4) Neural Network:

A method of artificial intelligence that tells computers how to process data in a way that is similar to the way the human brain works.

5) K- Nearest Neighbour:

To categorize the data, this algorithm essentially draws a fictitious boundary. The algorithm will attempt to forecast new data points to within one boundary line of the old ones.

```
from sklearn.tree import
DecisionTreeRegressor
from sklearn.svm import SVR
from sklearn.neural_network
import MLPRegressor
from sklearn.ensemble import
RandomForestRegressor
from sklearn.neighbors import
KNeighborsRegressor
```

Fig. 2. Code Snippet

IV. RESULTS

The algorithms' scores for accuracy are displayed below, along with a comparison of their performances.

TABLE I
COMPARISON TABLE

Serial No.	Result		
	Algorithm Name	*Accuracy*	*RMSE*
1	Support Vector Regressor	-1.198	68.7
2	Random Forest	98.323	24.5
3	Decision Tree	100	27.6
4	K- Nearest Neighbour	95.844	14.1
5	Neural Network	0.206	68.6

V. CONCLUSION

The regression method that performs the best is Random Forest with 0.983 efficiency and Decision Tree with 1.00 efficiency. Support Vector Machine is the regressor that has the lowest accuracy rate, which is -0.0119 accuracy rate. Other machine learning techniques, such as time series, clustering, association rules, and other ensemble approaches, can be added to this research to make it more comprehensive. To forecast rain for the hour after that, a similar procedure can be applied to hourly data using different parameters. A study may also be built using more thorough monitoring for a certain area or region, and constructing this type of model on a big data framework in order to enhance computation rate with higher precision and accuracy.

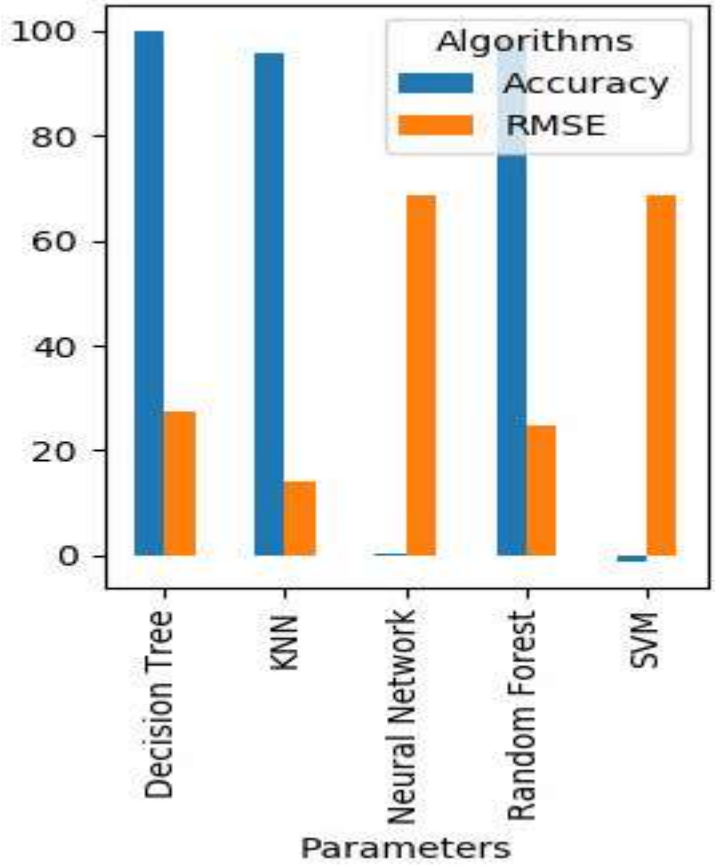

Fig. 3. Accuracy Graph

REFERENCES

[1] Aftab, S., Ahmad, M., Hameed, N., Bashir, M. S., Ali, I., Nawaz, Z. (2018). Rainfall prediction using data mining techniques: A systematic literature review. International Journal of Advanced Computer Science and Applications, 9(5), 143–150.

[2] Neelam Mishra, Hemanth Kumar Soni, Sanjiv Sharma, A. K. Upadhyay, ”Development and Analysis of Artificial Neural Network Models for Rainfall Prediction Using Time-Series Data”, Intelligent systems and Applications (I.J), 2018, 1, 16-23.

[3] Deepak Ranjan Nayak, Amitav Mahapatra, Pranati Mishra, “A Survey on Rainfall Prediction using Artificial Neural Networks”, International Journal of Computer Applications, volume 72- No.16, June 2013.

[4] Ali Haidar and Brijesh Verma. (2018). ”Monthly rainfall forecasting using a one-dimensional deep convolutional neural network.” IEEE
Access 6, pp. 69053-69063.

[5] Shreekanth Parashar, Tanveer Hurra, “A Study of Rainfall Using different Data Mining Techniques”, Research Gate, Article-May 2020.

[6] Narayan, V., & Daniel, A. K. (2022). FBCHS: Fuzzy Based Cluster Head Selection Protocol to Enhance Network Lifetime of WSN. ADCAIJ: Advances in Distributed Computing and Artificial

Intelligence Journal, 11(3), 285-307.

[7] Narayan, V., & Daniel, A. K. (2019). Novel protocol for detection and optimization of overlapping coverage in wireless sensor networks. Int. J. Eng. Adv. Technol, 8.

[8] Awasthi, S., Srivastava, A. P., Srivastava, S., & Narayan, V. (2019, April). A Comparative Study of Various CAPTCHA Methods for Securing Web Pages. In 2019 International Conference on Automation, Computational and Technology Management (ICACTM) (pp. 217-223). IEEE.

[9] Narayan, V., & Daniel, A. K. (2022). Energy Efficient Protocol for Lifetime Prediction of Wireless Sensor Network using Multivariate Polynomial Regression Model. Journal of Scientific & Industrial Research, 81(12), 1297-1309.

[10] Choudhary, S., Narayan, V., Faiz, M., & Pramanik, S. (2022). Fuzzy approach-based stable energy-efficient AODV routing protocol in mobile ad hoc networks. In Software Defined Networking for Ad Hoc Networks (pp. 125-139). Cham: Springer International Publishing.

[11] Srivastava, S., & Sharma, S. (2019, January). Analysis of cyber related issues by implementing data mining Algorithm. In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 606-610). IEEE.

[12] Srivastava, S., Yadav, R. K., Narayan, V., & Mall, P. K. (2022). An Ensemble Learning Approach For Chronic Kidney Disease Classification. Journal of Pharmaceutical Negative Results, 2401-2409.

[13] Srivastava, S., & Singh, P. K. (2022). Proof of Optimality based on Greedy Algorithm for Offline Cache Replacement Algorithm. International Journal of Next-Generation Computing, 13(3).

[14] Salagrama, S., Kumar, H. H., Nikitha, R., Prasanna, G., Sharma, K., & Awasthi, S. (2022, May). Real time social distance detection using Deep Learning. In 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES) (pp. 541-544). IEEE.

[15] Mahadani, A. K., Awasthi, S., Sanyal, G., Bhattacharjee, P., & Pippal, S. (2022). Indel-K2P: a modified Kimura 2 Parameters (K2P) model to incorporate insertion and deletion (Indel) information in phylogenetic analysis. Cyber-Physical Systems, 8(1), 32-44.

[16] Tyagi, N., Rana, A., Awasthi, S., & Tyagi, L. K. (2022). Data Science: Concern for Credit Card Scam with Artificial Intelligence. In Cyber Security in Intelligent Computing and Communications (pp. 115-128). Singapore: Springer Singapore.

[17] Awasthi, S., Kumar, N., & Srivastava, P. K. (2021). An epidemic model to analyze the dynamics of malware propagation in rechargeable wireless sensor network. Journal of Discrete Mathematical Sciences and Cryptography, 24(5), 1529-1543.

[18] Singh, M. K., Rishi, O. P., Awasthi, S., Srivastava, A. P., & Wadhwa, S. (2020, January). Classification and Comparison of Web Recommendation Systems used in Online Business. In 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 471-480). IEEE.

[19] Nawaraj Paudel, Tekendra Nath Yogi, "Comparative study of Machine Learning Algorithms for Rainfall Prediction- A case study in Nepal",International Journal of Advanced Research in Engineering and Technology (IJARET), Volume 11, Issue 10, 2020.