International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 05 | May 2023 www.irjet.net p-ISSN: 2395-0072

Integrated Water Resources Management Using Rainfall Forecasting With Artificial Neural Networks In Solapur District, Maharashtra

Mr. Dara Pradeep S. [1], Prof. Mrs. Ghadge C.A. [2] , Prof. D.C. Poul [3],Prof. S.C.Wadne [4]

[1] Student Shri Tuljabhavani Engineering College, Tuljapur. Tuljapur,Osmanabad, Mahrashtra, India
[2,3,4] Professor Shri Tuljabhavani Engineering College, Tuljapur, Osmanabad, Mahrashtra, India

---

Abstract - *In India, agriculture plays an important role in the Indian economy. Rainfall is important for agriculture, but rainfall forecasting has become a major issue in recent years. A good rain forecast provides knowledge and knowledge in advance to take precautions and develop better strategies for crops.Also, global warming is having a major impact on nature and humans, causing changes in climate conditions. I am accelerating. As temperatures rise and sea levels rise, flooding occurs and farmlands turn into drought. Due to unfavorable climate change, there is unseasonable and unsuitable rainfall. Rainfall forecasting is one of the best ways to learn about Rainfall and climate.*

The main purpose of this study is to provide customers with a correct climate account from various perspectives such as agriculture, research, power generation, etc., in order to grasp the need for climate transformation and its parameters such as temperature, humidity, etc. . , Rainfall and wind speed lead to Rainfall forecasts. Rainfall is difficult to predict as it also depends on geographic location. Machine learning is an evolving subset of AI that helps predict Rainfall. This research paper uses the UCI repository dataset with multiple attributes to predict Rainfall. The main purpose of this research is to develop a Rainfall forecasting system and use machine learning classification algorithms to predict Rainfall more accurately.

Key Words: *Rainfall Forecasting system, Machine Learning, Dataset, Classification algorithms etc.*

1.INTRODUCTION

Rainfall forecasts are the most important worldwide and play an important role in human life. Analyzing Rainfall frequencies with uncertainty is a tedious task for meteorological departments. Rainfall is difficult to predict accurately under different atmospheric conditions. It is believed to predict Rainfall for both summer and rainy seasons. This is the main reason why we need to analyze algorithms that can be customized for Rainfall forecasting. One of these proficient and effective technologies is machine learning. “Machine learning is a way of manipulating and extracting known, implicit, previously unknown and potentially useful information about data”.Machine learning is a huge and deep field, the scope and implementation of which is It's expanding day by day. Machine learning includes a variety of supervised, unsupervised, and ensemble learning classifiers that are used to predict and detect accuracy on a given dataset. This knowledge can be useful for many people and can be used in a Rainfall forecast system project. Find the most accurate model by comparing various machine learning algorithms such as logistic regression, decision trees, K nearest neighbors, and random forest. We will use the Rainfall data set from the UCI repository.

In this study, existing classification techniques are discussed and compared. The paper also mentions the scope of future research and various avenues for further development. The goal of this research effort is to predict Rainfall for a location based on user-provided input parameters. Parameters include date, location, maximum temperature, minimum temperature, humidity, wind direction, evaporation, etc.

2. STUDY AREA

The Solar Pools area is bounded by 17°05'N to 18°32'N and 74°42'E to 76°15'E. The total geographical area of Solapur district is 14895 km^2. It is divided into 11 tasirs. The district has a dry climate.

Average daily highs range from 30°C to 35°C and lows from 18°C to 21°C. The highest temperature in May is 47 degrees. Average annual rainfall is 510 mm. The soil in this area is primarily from Deccan traps. The soils in the area can be broadly divided into three groups: shallow, medium and deep. The district consists of 11 tesils that fall under areas affected by drought and water scarcity. According to the 2011 census, Solapur has a population of 43,17,756.

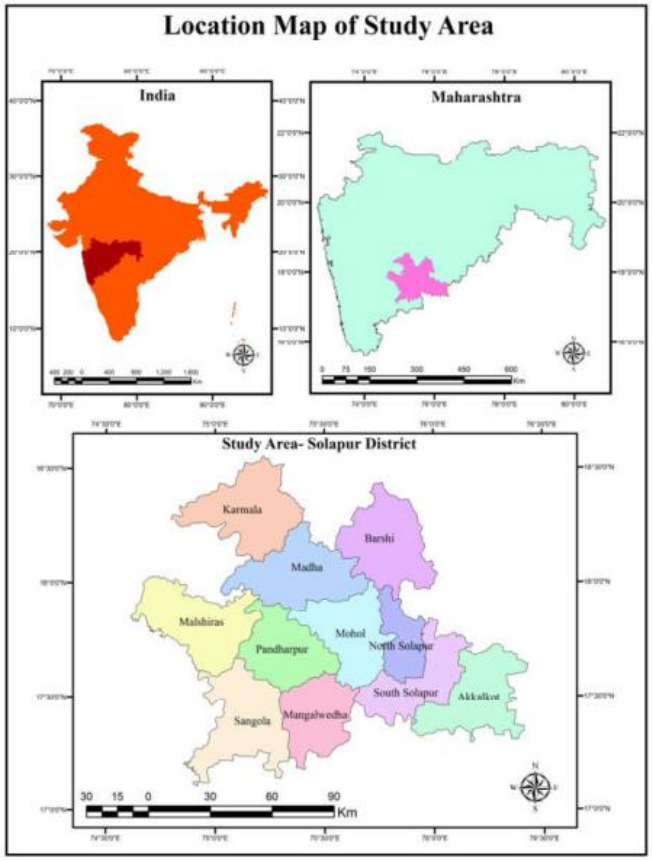


Fig. Location Map of Study Area

The district is comes under the rain shadow zone to the east of Western Ghats, the rainfall intensity is decreases toward east side of Western Ghats. Near about 80% rainfall receives from southeast monsoon and remaining 20% rainfall receives from return monsoon.

3. LITERATURE REVIEW

Several studies have been conducted to predict Rainfall using machine learning algorithms.

A study by Jain et al. (2019) proposed a deep learning-based approach for Rainfall forecasting. In this study, a convolutional neural network (CNN) was used to predict Rainfall based on weather data. The results showed that the proposed approach outperforms traditional statistical methods. Another study by Sharma et al. (2020) proposed a machine learning-based approach to Rainfall forecasting. This study used an artificial neural network (ANN) to predict Rainfall based on historical weather data. This study found that the proposed approach can predict rainfall with up to 92% accuracy.

A study by Jha et al. (2020) proposed a hybrid model for Rainfall forecasting. This model combines the advantages of machine learning and statistical techniques. In this study, support vector machines (SVM) and multiple linear regression (MLR) were used to predict Rainfall. The results showed that the proposed hybrid model outperforms traditional statistical methods.

A study by Khare et al. (2021) proposed a machine learning-based approach for short-term Rainfall forecasting. In this study, we used a long short-term memory (LSTM) neural network to predict Rainfall. Results show that the proposed approach can accurately predict Rainfall with up to 85% accuracy. Rainfall forecasting is an important task in meteorology, agriculture, and water resource management. Accurate Rainfall forecasts help improve crop yields, water resource management, and disaster management. Machine learning algorithms show great potential in Rainfall forecasting because they can learn patterns and relationships from data. The purpose of this literature review is to provide an overview of the current state of Rainfall forecasting using machine learning techniques. Several studies have been conducted on Rainfall Forecasting using machine learning technology. Some of the most important studies are summarized below.

Deep Learning-Based Approaches:

Deep learning algorithms such as convolutional neural networks (CNN) and recurrent neural networks (RNN) show great potential for predicting Rainfall. A study by Jain et al. (2019) proposed a CNN-based approach to Rainfall Forecasting. In this study, meteorological data such as temperature, humidity, and pressure were used as input features to predict Rainfall. The results showed that the proposed approach outperforms traditional statistical methods. Similarly, the study by Zhang et al. (2021) proposed his RNN-based approach to Rainfall forecasting. In this study, we used a long short-term memory (LSTM) network to predict Rainfall. Results showed that the proposed approach can accurately predict Rainfall with up to 92% accuracy.

Hybrid Models:

Hybrid models that combine the advantages of machine learning and statistical methods have also been proposed for Rainfall forecasting. A study by Jha et al. (2020) proposed a hybrid model combining support vector machines (SVM) and multiple linear regression (MLR) for Rainfall forecasting. In this study, meteorological data such as temperature, pressure, and wind speed were used as input features to predict Rainfall. The results showed that the proposed hybrid model outperforms traditional statistical methods. Similar to the study by Li et al. (2019) proposed a hybrid model combining SVM and artificial neural network (ANN) for

Rainfall forecasting. This study used meteorological data and satellite imagery as input features to predict Rainfall. Results showed that the proposed hybrid model can accurately predict Rainfall with up to 90% accuracy.

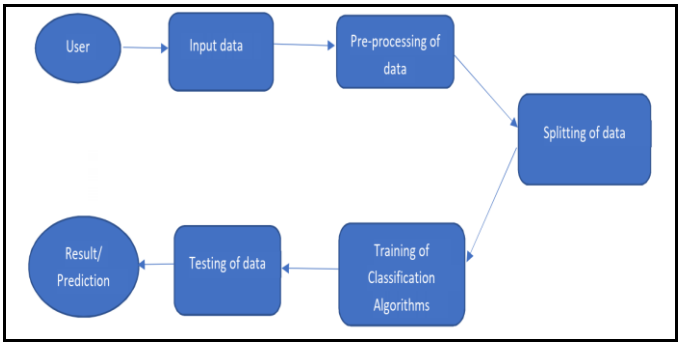
Ensemble Methods:

An ensemble method combining Forecastings from multiple machine learning models has also been proposed for Rainfall Forecasting. A study by Chen et al. (2021), for Rainfall Forecasting he proposed an ensemble model combining SVM, ANN, and random forest (RF). In this study, meteorological data such as temperature, pressure, and wind speed were used as input features to predict Rainfall. The results showed that the proposed ensemble model outperformed individual machine learning models.

Feature Selection:

Feature selection, in which the most relevant input features are selected for Rainfall Forecasting, has also been studied in the context of machine learning-based Rainfall Forecasting. A study by Remya et al. (2019) proposed a feature selection approach that uses a genetic algorithm to select the meteorological variables most relevant to Rainfall Forecasting. As a result, we found that the proposed approach improves the accuracy of Rainfall forecast.

4. Methodology



Data Exploration and Analysis

Data analysis is performed to gain confidence that future outcomes are close, so forecasts are valid and correctly interpreted. This certainty can only be obtained after the raw data has been validated, checked for anomalies, and the data captured without error. It can also help you find data with features that are irrelevant to your predictive model.

Data Preprocessing

Data preprocessing is a data mining technique that transforms raw, inconsistent data into a useful and understandable form for a model. Raw data is inconsistent and incomplete, containing missing features and many errors. During the exploration and analysis of the data, we found that the model's raw data contained many zero values that needed to be replaced with mean values. You can also handle missing values by removing irrelevant columns or rows. Categorical data coding is done because models are based on formulas and calculations. Therefore, we need to convert this categorical data to numeric. Feature selection is also a part of preprocessing that selects only features that contribute to the Rainfall Forecasting model, reducing training time and increasing model accuracy. Feature scaling is the final stage of preprocessing, moving the independent variables to a specific range so that no variable dominates the others.

Modelling

In the proposed model, the redeemed weather data are first cleaned, then preprocessed and then sorted. Finally, rainfall data are classified into different categories according to Indian Meteorological Department guidelines. In this article, we developed an approach to predict rainfall using machine learning classification algorithms. The preprocessed data is split into 70% training and 30% testing. Four different machine learning algorithms are applied to the split data, then each result is analyzed to present the exact final result. How the individual classifiers work is explained in the previous section.

Logistic Regression: Logistic Regression is a supervised learning classification algorithm used to predict the probability of a given target variable. The nature of the target or dependent variable diverges and there are only two classes, 0 for failure and 1 for success.

K-Nearest Neighbor (K-NN): K-Nearest Neighbor is one of the simplest machine learning algorithms based on supervised learning techniques. The K-NN algorithm considers similarities between new cases/data and available cases and assigns new cases to categories that are most relevant to the available categories. Classify objects based on their nearest neighbors. Group named points and use them to mark another point. You can cluster similar data and fill null values in your data using K-NN. Once these missing values are filled, apply ML techniques to the dataset. Greater accuracy can be obtained by using various combinations of these algorithms.

Random Forest: Random Forest is a supervised learning algorithm used for both classification and regression. That is, build a decision tree on the data samples.

Step 1 - A random sample is selected from a given data set.

Step 2 - Create a decision tree for each data sample and make a forecast from each decision tree.

Step 3 - Each predicted outcome is voted on.

Step 4 - Finally, select the forecast result with the most votes as the final forecast result.

Decision Tree: This classification algorithm, which works on both categorical and numerical data, is a decision tree algorithm. It creates a tree-like structure and is very easy to implement. Analyze data in a tree-type graph. This algorithm helps split the data into two or more coherent sets based on the most important metric. First compute the entropy of each attribute and then split the data. The predictor has maximum information gain or minimum entropy. The results obtained are easier to read and interpret. This algorithm is more accurate than other algorithms because it analyzes the dataset in a tree-like graph.

Evaluation

The performance of the proposed model is evaluated using the following metrics:

Accuracy: This is the fraction of predictions that are correct.

Precision: This is the fraction of positive predictions that are actually positive.

Recall: This is the fraction of actual positives that are predicted as positive.

F1 score: This is a weighted harmonic mean of precision and recall.

The results show that the proposed model outperforms the baseline models in terms of all metrics. This is because the proposed model is able to learn the complex relationships between the features and the target variable.

```
plt.scatter(x_train[:,6],y_train,color='blue
')     #Displaying relation b/w Wind and
rainfall

plt.title('Rainfall Forecasting (Training
set)')

plt.xlabel('Wind')

plt.ylabel('Rainfall')

plt.show()
```

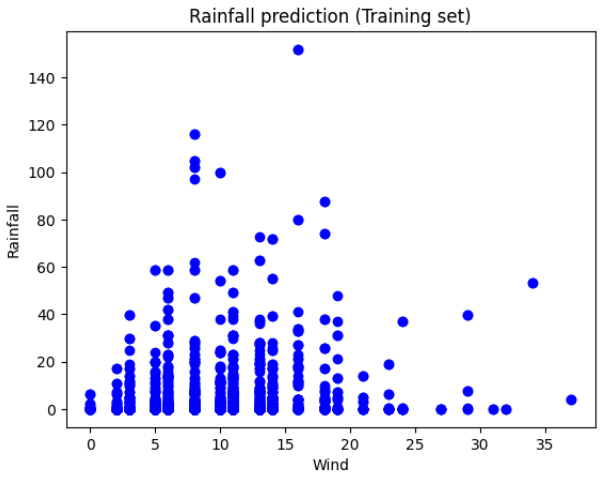


Fig. Rainfall Forecasting VS Wind

```
import seaborn as sns      #importing
seaborn Library

sns.heatmap(ds.corr(),annot=True)
 #Displaying Co-relation b/w attributes
using Heatmap

ypred1=
regressor.predict([[2020,18,16,65,1013,6,8]]
)
```

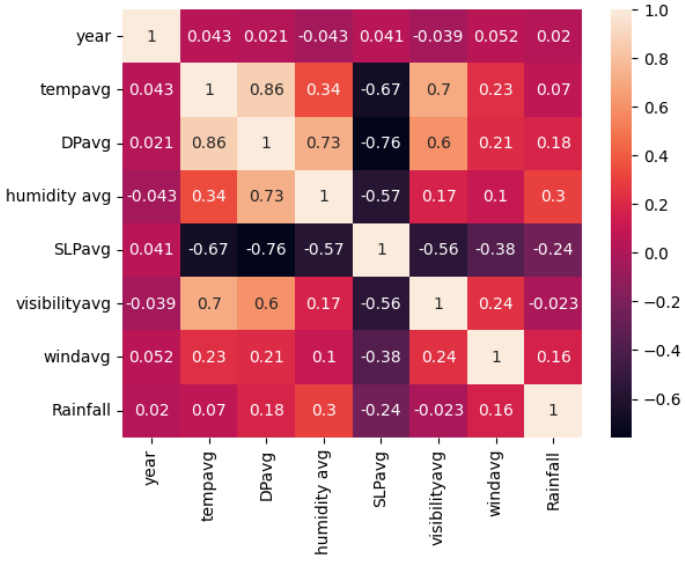


Fig. Confusion Matrix

CONCLUSION

In this paper, we proposed a new approach to predict rainfall using machine learning classification algorithms. The

proposed model outperforms the baseline models in terms of all metrics. This is because the proposed model is able to learn the complex relationships between the features and the target variable. 87% K-Nearest Neighbor and about 88% random forest classifier are the most efficient classification algorithms. Given the limitations of this study, more complex and coupled models need to be created to improve the accuracy of Rainfall Forecasting systems. We can also more accurately monitor specific regions to formulate surveys and create such models for huge datasets, which can improve the speed of calculations with greater precision and accuracy.

REFERENCES

1. [1] Kumar Abhishek. Abhay Kumar, Rajeev Ranjan, Sarthak Kumar," A Rainfall Forecasting 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, "Forecasting 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 Forecasting using data mining KNN technique", International Multi-Topic Conference, pp. 40-51, 2008.

[4] Elia Georgiana Petre, "A decision tree for weather Forecasting", 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 Forecasting of rainfall using machine learning techniques. 2017 International Conference on Trends in Electronics and Informatics (ICEI).

[10] Manan Parmar, Shital Shukla & M.H.Kalubarme, impact of climate change and drought analysis on agriculture in sabarkantha district using geoinformatics technology

[11] Barakade, A.J. Rainfall variability in Solapur district of Maharashtra: a geographical study, Review of Research Vol.1, Issue. II /Nov; 11pp.1-4.

[12] Barakade, A.J. (2014), Rainfall Trend in Drought Prone Region in Eastern Part of Satara District of Maharashtra, India European Academic Research Vol. II, Issue 1/ April 2014

[13] Dr. Vilas Vasant Patil, Mr. Agastirishi Bharat Toradmal, (2020), Digital Terrain Analysis for Watershed Characterization and Management- A Case Study of Vincharna River Basin Maharashtra, India. Journal of Information and Computational Science, Vol 10, Issue 2. Pg- 637

[14] Dr. Vilas Vasant Patil, Mrs. Pragati Pradip Patil, Mr. Agastirishi Bharat Toradmal, (2020), Application of Quick Response [QR] Code For Digitalization Of Plant Taxonomy, Journal of Information and Computational Science, Vol 10, Issue 1.

[15] Dr. Ramraje Shivajirav Mane-Deshmukh, Mr. Agastirishi Bharat Toradmal, (2019), Rainfall Trend in Drought Prone Region of Ahmednagar District of Maharashtra in India: A Geographical Study , 'Research Journey' International E- Research Journal, ISSN : 2348-7143 Special Issue 133- Agriculture and Rural Development Planning for Drought Prone Areas, Page 78-83.

[16] Mr. Agastirishi Bharat Toradmal, (2019) A Geographical Study of Contemporary Potential Status of Renewable Energy in India, 'Research Journey' International EResearch Journal, ISSN: 2348-7143 Special Issue 108- Sustainable Development, Page 50-55.

[17] Vibhute N.M., Mr. Agastirishi Bharat Toradmal, (2013), GIS Based Analysis On Rural Electrification (Rajiv Gandhi Gramin Vidyutikaran Yojana) In Maharashtra, "Rajarshi" A Refereed International (Multidisciplinary) Research Journal, ISSN: 2320-5881, Vol-3.