



International Journal of Research in Agronomy

E-ISSN: 2618-0618
P-ISSN: 2618-060X
© Agronomy
NAAS Rating (2026): 5.20
www.agronomyjournals.com
2026; SP-9(5): 394-398
Received: 22-02-2026
Accepted: 30-03-2026

Devika Santhosh
M.Sc. Research Scholar,
Department of Agricultural
Statistics and Social Science (L.),
College of Agriculture, IGKV,
Raipur, Chhattisgarh, India

Dr. Lakshmi Narsimhaiah
Assistant Professor, Department of
Agricultural Statistics and Social
Science (L.), College of Agriculture,
IGKV, Raipur, Chhattisgarh,
India

Vijayakumar S
M.Sc. Research Scholar,
Department of Agricultural
Statistics and Social Science (L.),
College of Agriculture, IGKV,
Raipur, Chhattisgarh, India

Dr. Girish Chandel
Vice-Chancellor, College of
Agriculture, IGKV, Raipur,
Chhattisgarh, India

Corresponding Author:
Devika Santhosh
M.Sc. Research Scholar,
Department of Agricultural
Statistics and Social Science (L.),
College of Agriculture, IGKV,
Raipur, Chhattisgarh, India

Monthly rainfall prediction using machine learning models: A comparative study of decision tree, SVM, and KNN in Raipur, Durg, and Bilaspur Districts of Chhattisgarh

Devika Santhosh, Lakshmi Narsimhaiah, Vijayakumar S and Girish Chandel

DOI: <https://www.doi.org/10.33545/2618060X.2026.v9.i5Se.5598>

Abstract

Accurate predictions of rainfall are crucial for managing agriculture and water resources in the central plains of Chhattisgarh, an area where the productivity of crops is significantly influenced by the variability of the monsoon. This research assesses the efficacy of three machine learning models—Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)—for forecasting monthly rainfall in the districts of Raipur, Bilaspur, and Durg. Historical monthly weather data from 1981 to 2022, obtained from Nasa power, were processed and utilized for training and validating the models. The performance of the models was evaluated using metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), the coefficient of determination (R^2), and Mean Square Error. The results reveal that the KNN model consistently excelled over DT and SVM in all districts, achieving R^2 values of 0.81 for Raipur, 0.86 for Bilaspur, and 0.79 for Durg, along with lower RMSE and MAE figures. Due to KNN's outstanding performance, forecasts for rainfall from 2023 to 2028 were generated using the monthly rainfall time series data spanning from 1981 to 2022, highlighting district-specific trends that can support localized agricultural decision-making. The results illustrate that KNN is an effective and dependable approach for district-level rainfall prediction within the region.

Keywords: Rainfall prediction, Chhattisgarh, Raipur, Bilaspur, Durg, machine learning, KNN, support vector machine, decision tree

Introduction

Precipitation is a crucial weather variable that significantly impacts various essential areas, including agriculture, water resource management, and hydroelectric power generation. Agricultural practices in developing nations heavily rely on rainfall. Chhattisgarh, situated in central India, is a state primarily focused on agriculture, where the monsoon rains significantly influence crop yields and the livelihoods of rural communities. In this state, the central plains—especially the districts of Raipur, Bilaspur, and Durg—play a key role in agricultural production, particularly in the cultivation of rice.

Traditional methods for forecasting, such as various weather prediction models and statistical approaches like ARIMA, SARIMA, and regression models, have been extensively utilized for predicting rainfall. Nevertheless, these methods often encounter difficulties due to the complex, nonlinear, and ever-changing characteristics of weather systems. Recent progress in machine learning and artificial intelligence has exhibited notable potential in enhancing the precision of rainfall forecasting by identifying complex patterns within meteorological data (Shukla *et al.*, 2024) ^[1]. Research has indicated that machine learning models can effectively forecast rainfall, thereby boosting agricultural yield and ensuring the availability of food and water supplies (Wani, O.A *et al.*, 2023) ^[2]. A machine learning algorithm functions by recognizing patterns and relationships within data to generate predictions or make decisions without the need for explicit programming for each specific task. The incorporation of ML models like Decision Trees, KNN, Random Forest, Support Vector Machines (SVM), and XGBoost has shown considerable

promise in minimizing forecasting errors and improving reliability (Patel *et al.*, 2023) [3].

The current research aims to predict monthly rainfall in the districts of Raipur, Bilaspur, and Durg located in the central plains of Chhattisgarh by utilizing weather parameters such as wind speed, relative humidity, temperature, Dew, Surface pressure, Minimum and Maximum temperature. Weather parameters play a vital role in predicting rainfall as they impact the atmospheric processes that lead to precipitation. Temperature influences air pressure, humidity, and evaporation, which are critical for cloud formation. Humidity indicates the amount of moisture present in the air, heightening the likelihood of condensation and rainfall. Variations in air pressure assist in forecasting weather trends, while wind speed and direction are essential for transporting moisture and influencing storm formation. An examination is conducted on three machine learning algorithms: Decision Trees, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), to identify which model provides the most accurate predictions of rainfall based on these factors. After identifying the most effective model, it will be employed to estimate future rainfall using historical monthly rainfall data from these districts.

The machine learning models' accuracy is assessed using R-squared (R^2), Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE). The model that demonstrates the highest R^2 value and the lowest error metrics will be deemed the most effective for predicting and forecasting rainfall.

Study Area

The research examines three districts—Raipur, Bilaspur, and Durg—situated in the central plains of Chhattisgarh, India. This area is vital to the state's agriculture and economy, with monsoon rainfall serving as a key element for crop production and water supply. Raipur, the state capital located centrally in Chhattisgarh, typically experiences moderate to high levels of monsoon rainfall that is distributed fairly evenly during the rainy season. In contrast, Bilaspur, found in the northern plains, exhibits greater variability in rainfall from year to year, impacting the timing and intensity of rainfall events, thus complicating agricultural planning. Durg, located in the western plains, usually obtains stable annual rainfall, though there are occasional fluctuations from long-term averages due to regional climatic influences. The rainfall in Durg supports a variety of cropping systems that rely on consistent moisture availability. Together, these three districts illustrate a range of rainfall patterns within the central plains of Chhattisgarh, making them ideal for comparative studies on rainfall prediction and forecasting utilizing machine learning models.

Materials and Methods

Data collection

The research will rely on the historical meteorological data. The weather data for variables such as relative humidity, rainfall, maximum temperature, minimum temperature, and wind speed will be collected from:

1. NASA POWER (<https://power.larc.nasa.gov/data-access-viewer/>)
2. IMD government site (<https://mausam.imd.gov.in>)

Data Preprocessing

The term data preprocessing is frequently used in the area of machine learning to refer to the processes involved in organizing, cleaning, and preparing raw data including detection

and removal of outliers before using it to build machine learning models. Then the whole dataset is divided into two parts, the training phase and the testing phase

Analytical Methods

All the analysis will be carried out using open source programming languages like R or Python. This platform provides strong features for data preprocessing, model training, performance evaluation, and visualization. In the present investigation machine learning models, namely, Decision trees, SVM and KNN were applied for annual average rainfall prediction.

Decision trees

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.

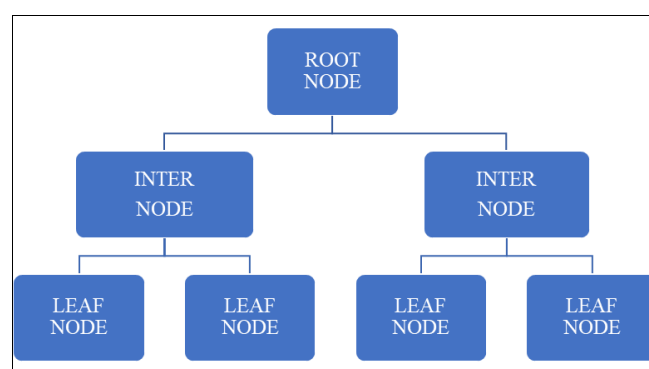


Fig 1: Nodes of Decision trees.

As you can notice from the Figure 1, a decision tree begins with a root node, which has no incoming branches. The outgoing branches of the root node then pass to the internal nodes, or decision nodes. Depending on the features available, both types of nodes make evaluations to create homogenous subsets, which are represented by leaf nodes, or terminal nodes. The leaf nodes are all the possible outcomes in the dataset (IBM, 2024) [6].

KNN

The k-nearest neighbours (KNN) algorithm is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. It is one of the popular and simplest classification and regression classifiers used in machine learning today.

In the present study KNN makes an annual rainfall prediction by identifying similar past weather data to an input and using those to estimate the amount of rainfall. The model initially needs historical data with attributes like temperature, humidity, wind speed, atmospheric pressure, and past rainfall. In making a prediction, KNN computes the distance from the new data point to all the previous data points, often using Euclidean distance.

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_j - X_{ij})^2}$$

where,

$d(x,y)$ = Euclidean distance

x = test samples. ($x_1, x_2, \dots, x_d$)

X_i = Historical observations ($X_{i1}, X_{i2}, \dots, X_{id}$).

j =Different weather parameters.

It then picks the K nearest data points and calculates the predicted rainfall by taking their rainfall values' average in regression or the most frequent class in classification. As KNN is a non-parametric model, it does not assume anything about the distribution of the data, thus being useful in modelling complex, non-linear weather patterns.

Support Vector Machine

The key idea behind the SVM algorithm is to find the hyperplane that best separates two classes by maximizing the margin between them. This margin is the distance from the hyperplane to the nearest data points (support vectors) on each side. The best hyperplane, also known as the “hard margin,” is the one that maximizes the distance between the hyperplane and the nearest data points from both classes. This ensures a clear separation between the classes. Support Vector Machine (SVM) is a powerful classification algorithm that efficiently deals with linearly and nonlinearly separable data. It attempts to determine the best margin that linearly separates classes optimally. If the data is linearly separable, a linear SVM identifies the best separating hyperplane, while for nonlinearly separable data, SVM applies the kernel trick to map the data to a higher-dimensional space, where it can then be separated by a hyperplane (IBM, 2024) [6].

Evaluation criteria

To compare the model accuracy the following given in Table 1 evaluation criteria have been used:

Table 1: Evaluation Criteria used to determine the best model

S. No.	Evaluation Criteria	Formula
1	Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$
2	Mean squared Error (MSE)	$MSE = \frac{1}{N} \sum_{i=1}^N y_i - \hat{y}_i ^2$
3	Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N y_i^2}$
4	Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$

y_i and $\hat{y}_i$ are the actual value and predicted value of the response variable and N is the number of data points. The best-performing model was then applied to forecast monthly rainfall for Raipur, Bilaspur, and Durg for the years 2023 to 2028 using historical rainfall time series data as input.

Results and Discussion

The machine learning models — Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbours (KNN) — were applied to predict monthly rainfall for the selected districts. The performance of each model was evaluated using four

metrics: Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE). The evaluation results for the three selected districts are summarised in Table 2.

Table 2: Performance comparison of Decision Tree, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) models for monthly rainfall prediction in Raipur, Bilaspur, and Durg districts of Chhattisgarh, evaluated using R^2 , RMSE, MSE, and MAE.

District	Model	R^2	RMSE	MSE	MAE
Raipur	Decision Tree	0.62	81.67	6670.50	44.27
	KNN	0.81	57.70	3330.02	33.95
	SVM	0.75	65.18	4249.35	44.64
Bilaspur	Decision Tree	0.68	59.14	3498.34	35.44
	KNN	0.86	47.34	2241.19	31.14
	SVM	0.78	58.51	3423.87	44.44
Durg	Decision Tree	0.64	83.09	6904.44	48.40
	KNN	0.79	62.99	3957.52	38.62
	SVM	0.74	70.70	4999.51	46.90

In Raipur, the R^2 values showed that KNN (0.81) delivered the best predictive accuracy, followed by SVM (0.75) and DT (0.62). The RMSE values were the lowest for KNN (57.70 mm), indicating superior prediction performance compared to SVM (65.18 mm) and DT (81.67 mm). This pattern was also evident for MSE and MAE, where KNN consistently generated lower error values. In Bilaspur, the KNN model again surpassed the other models, achieving the highest R^2 (0.86) and the lowest RMSE (47.34 mm), MSE (2241.19 mm²), and MAE (31.14 mm). SVM showed moderate performance with an R^2 of 0.78, while DT displayed relatively lower accuracy (R^2 = 0.68). For Durg, KNN obtained the best results with an R^2 of 0.79, an RMSE of 62.99 mm, an MSE of 3957.52 mm², and an MAE of 38.62 mm. SVM and DT exhibited slightly less predictive power, with DT registering the highest error values among the three models. In summary, KNN consistently demonstrated superior prediction performance across all three districts, evidenced by higher R^2 and lower RMSE, MSE, and MAE values, confirming its strength and dependability for forecasting monthly rainfall using weather parameters. In light of these results, KNN was chosen as the optimal model for continued forecasting with rainfall time series data. Based on the K-Nearest Neighbors (KNN) model, which was identified as the best-performing algorithm for rainfall prediction in the Central Plains of Chhattisgarh, the average monthly rainfall for the forecast period (2023–2028) was estimated for Raipur, Bilaspur, and Durg districts. The model utilized historical monthly rainfall data from 1991 to 2022 to generate future projections. The results indicate year-to-year variability in rainfall, reflecting the inherent climatic fluctuations in the region.

Table 3: Forecasted Monthly Average Rainfall (mm) in Raipur District (2023–2028)

Month	2023	2024	2025	2026	2027	2028
January	19.10	22.80	21.10	14.00	13.50	29.60
February	10.60	17.90	9.90	15.80	16.90	17.60
March	16.00	10.50	11.60	5.50	8.70	7.90
April	12.70	19.30	6.70	23.10	18.80	20.40
May	22.10	27.60	44.40	17.20	26.50	29.10
June	150.60	267.70	133.00	183.60	205.50	190.40
July	306.70	228.50	344.80	273.10	291.20	285.60
August	233.30	314.50	298.80	295.50	296.20	306.80
September	171.80	224.70	151.10	241.70	188.20	208.40
October	50.80	47.30	32.60	58.60	49.40	58.60
November	8.50	10.50	5.10	0.10	0.50	0.50
December	4.50	10.40	12.90	3.70	3.90	3.90
Total Monthly average Rainfall	1006.70	1201.70	1072.00	1131.90	1119.30	1158.80

Table 4: Forecasted Monthly Average Rainfall (mm) in Bilaspur District (2023–2028)

Month	2023	2024	2025	2026	2027	2028
January	12.80	19.80	37.50	14.10	40.10	14.20
February	13.80	21.60	21.90	28.40	21.10	30.30
March	8.20	16.10	11.20	15.80	11.20	15.80
April	12.70	16.40	13.00	24.50	13.00	19.90
May	30.20	28.60	37.90	18.90	37.90	14.10
June	185.80	209.90	114.10	228.80	114.10	223.60
July	342.60	317.60	370.70	290.90	388.20	304.30
August	243.50	285.80	311.90	238.80	332.10	277.80
September	185.10	199.70	184.30	211.00	184.30	250.10
October	39.60	23.70	44.40	23.70	44.40	53.50
November	8.00	10.40	4.00	10.40	4.10	2.10
December	12.60	7.40	15.20	7.40	12.80	1.90
Total Monthly average Rainfall	1094.90	1157.00	1166.10	1112.70	1203.30	1207.60

Table 5: Forecasted Monthly Average Rainfall (mm) in Durg District (2023–2028)

Month	2023	2024	2025	2026	2027	2028
January	15.30	24.20	11.60	9.10	23.30	11.40
February	12.70	12.90	8.30	17.60	10.00	15.30
March	10.80	4.40	5.70	4.70	5.80	7.30
April	13.10	40.30	28.20	31.90	16.60	32.80
May	33.10	18.90	22.60	21.30	24.70	19.40
June	204.50	134.20	105.40	266.20	156.40	171.00
July	403.50	314.90	345.00	269.90	344.00	279.40
August	245.50	340.80	304.50	300.40	308.80	305.30
September	192.90	166.10	225.70	272.40	172.20	206.20
October	43.80	45.00	70.20	74.50	47.00	59.50
November	9.80	6.40	19.80	0.60	2.10	5.90
December	11.90	10.00	0.20	3.10	10.00	3.70
Total Monthly average Rainfall	1196.90	1118.10	1147.20	1271.70	1120.90	1117.20

The forecast indicates significant seasonal and annual variability in rainfall for Raipur, Bilaspur, and Durg districts from 2023 to 2028. The majority of annual rainfall in all three districts occurs between June and September, with July and August consistently showing the highest monthly totals, underscoring their role as the primary monsoon months. Durg is projected to experience the highest peak rainfall in July, with amounts surpassing 400 mm in 2023. Bilaspur typically records a slightly greater total annual rainfall compared to Raipur and Durg, especially notable in the years 2027 and 2028. Raipur demonstrates relatively consistent rainfall patterns, with less year-to-year variation in total precipitation than the other two districts. 265Rainfall during the winter months (November to February) is expected to be minimal, generally remaining below 20 mm, aside from occasional spikes in January in Raipur and Bilaspur. The pre-monsoon months (April–May) are predicted to have moderate variability, featuring sporadic high-rainfall events in Durg (such as in April 2024) and Bilaspur (in May 2023).

Conclusion

This research examined the performance of Decision Tree, SVM, and KNN models in predicting monthly rainfall in the regions of Raipur, Bilaspur, and Durg. Among the models tested, KNN proved to be the most precise, obtaining the best R^2 value and the lowest error rates. By employing KNN, rainfall projections for 2023–2028 were made, indicating peak rainfall during July and August, with distinct variations across the districts, revealing that Bilaspur is expected to receive the most rain while Durg is forecasted to receive the least. These results underscore the effectiveness of KNN in dependable rainfall

forecasting and its relevance for agricultural planning and water resource management in Chhattisgarh.

Future Prospects

In the upcoming years, this research can be enhanced in various ways. More sophisticated models, such as hybrid machine learning approaches that leverage the advantages of different algorithms, could be utilized to improve the precision of rainfall forecasts. In addition to historical rainfall data, supplemental information like satellite imagery, temperature, humidity, soil moisture, and large-scale climate phenomena such as El Niño can be incorporated to enhance the reliability of the models.

The geographical scope of the study could also be broadened to encompass additional districts or even the entire state, which would provide a more comprehensive understanding of rainfall trends across Chhattisgarh. Implementing real-time data and automated forecasting systems will facilitate timely alerts for farmers and policymakers.

In agriculture, this advancement can assist farmers in determining the optimal times for planting and harvesting their crops. For the government, it could aid in improving water management, disaster preparedness, and strategic planning for droughts or flooding. In summary, refining and broadening this research can transform rainfall prediction into a valuable resource for both agricultural practices and resource management in the region.

References

- Shukla R, Kumar A, Singh P. Application of machine learning models in rainfall prediction over Indian monsoon regions. *Theoretical and Applied Climatology*. 2024;157(3):1345-1359. doi:10.1007/s00704-023-04325-1.
- Wani OA, Khan FA, Bhat MS. Evaluation of machine learning approaches for rainfall prediction: a case study of Indian states. *Environmental Monitoring and Assessment*. 2023;195(6):733-733. doi:10.1007/s10661-023-11378-7.
- Patel R, Jha A, Kumar S. Performance comparison of statistical and machine learning techniques for rainfall forecasting in India. *Modeling Earth Systems and Environment*. 2023;9(2):1451-1464. doi:10.1007/s40808-022-01557-y.
- NASA POWER. NASA Prediction of Worldwide Energy Resources (POWER) Data Access Viewer. 2024.
- India Meteorological Department. Mausam: weather information and data services. 2024.
- IBM. Decision tree and support vector machine explained. 2024.
- Breiman L, Friedman J, Olshen R, Stone C. Classification and regression trees. Belmont: Wadsworth International Group; 1984.
- Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;20(3):273-297. doi:10.1007/BF00994018.
- Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. 1967;13(1):21-27. doi:10.1109/TIT.1967.1053964.
- Ghosh S, Mujumdar PP. Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources*. 2008;31(1):132-146. doi:10.1016/j.advwatres.2007.07.005.
- Kannan S, Ghosh S. A nonparametric kernel regression model for downscaling precipitation to river basin in India. *Hydrology and Earth System Sciences*. 2013;17(3):887-905. doi:10.5194/hess-17-887-2013.
- Mishra AK, Desai VR. Drought forecasting using stochastic

- models. *Stochastic Environmental Research and Risk Assessment*. 2005;19(5):326-339. doi:10.1007/s00477-005-0238-4.
13. Bhowmik AK, Das D. Rainfall variability and trend analysis in Chhattisgarh. *Journal of Agrometeorology*. 2019;21(1):29-36. doi:10.54386/jam.v21i1.526.
 14. Chattopadhyay S, Chattopadhyay G. Identification of best predictor variables for Indian summer monsoon rainfall using statistical and neural network models. *Atmospheric Research*. 2010;98(2-4):281-289. doi:10.1016/j.atmosres.2010.07.011.
 15. Goyal MK, Ojha CSP. Evaluation of neural network models for runoff simulation. *Water Resources Management*. 2010;24(6):1065-1080. doi:10.1007/s11269-009-9489-9.
 16. Jain SK, Kumar V. Trend analysis of rainfall and temperature data for India. *Current Science*. 2012;102(1):37-49.
 17. Sen PN. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*. 1968;63(324):1379-1389. doi:10.1080/01621459.1968.10480934.
 18. Tripathi S, Srinivas VV, Nanjundiah RS. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology*. 2006;330(3-4):621-640. doi:10.1016/j.jhydrol.2006.04.030.
 19. Yadav SK, Singh R. Machine learning techniques for climate and weather prediction: a review. *Earth Science Informatics*. 2021;14(4):1687-1706. doi:10.1007/s12145-021-00663-2.
 20. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 2003;50(C):159-175. doi:10.1016/S0925-2312(01)00702-0.