[1]Ms.Nirmal Mungale

[2]Dr.Jayshri Shinde

Machine Learning Techniques to Predict Rainfall of Vidarbh Region

Journal of Electrical Systems

Abstract: - Accurate daily rainfall forecasting is vital for proper agricultural planning and managing resources efficiently. This study evaluates the performance of various methods, with an emphasis on temperature and precipitation, such as linear regression, random forest, SVM regression, and XGBoost. Linear regression displays minimal RMSE and a flawless R-squared, whereas random forest performs effectively with low RMSE and encouraging R2 results. However, the presence of a negative R-squared indicates potential overfitting. The MAE, MSE, and RMSE statistics for SVM are competitive. The study draws attention to the unexplored Vidarbha region, which includes 11 districts, using Nagpur district as a representative instance. Additionally, future plans involve the utilization of deep learning models like ARIMA and LSTM to enhance rainfall prediction accuracy across Vidarbha. This investigation yields valuable insights into climate prediction, offering support for well-informed decision-making.

Keywords: Machine learning, linear regression, Random forest, RF, XGBoost, Rainfall prediction

I. INTRODUCTION:

Linear regression, SVM, and XGBoost are among the ML approaches used for daily rainfall prediction. These approaches employ a variety of tactics to improve predicting outcomes [1]. Accurate rainfall forecasting is critical for increasing agricultural productivity, guaranteeing food security, and maintaining a regular water supply for a country's residents. Inadequate rainfall can have a negative influence on both aquatic ecosystems and water quality, eventually affecting agricultural production. Given the complicated relationship between agriculture and water quality and daily and yearly rainfall patterns, attaining exact estimates for daily rainfall is a challenging undertaking with far-reaching ramifications for agricultural and water resource management.

Temperature values ranging from 14.38 degrees Celsius to 32.58 degrees Celsius are included in the dataset, which covers January 1, 2016, to December 31, 2022. At the same time, humidity levels range from 49.62% to 98.55%. However, despite the dataset spanning 11 districts, the focus of this study is on the Nagpur district as an instructive example. To aid comparison, the remaining ten districts are shown in tabular form. Using data mining techniques [2], large-scale data analysis, and additional machine learning algorithms, researchers enhanced precision for precipitation on each day, monthly, and yearly basis projections. Based on the outcomes of the research, the method for forecasting has recently evolved from data analysis procedures to predictive machine learning techniques. For instance, researchers [3] observed that machine learning algorithms outperform classic deterministic methods for predicting weather and rainfall. As an outcome, this study investigated different algorithms for machine learning in order to discover which were most accurate for rainfall forecasting.

Rainfall frequency and intensity are influenced by a number of environmental variables [4]. Affecting factors include humidity levels, temperature, natural light, threats, condensation, and other factors on the presence and severity of rainfall, either directly or indirectly. As an outcome, this study's goal was to employ ML methods to recognize the key weather factors that cause rainfall and anticipate the intensity of typical precipitation [5]. Nagpur csv is the dataset, which has been preprocessed to make it appropriate for the experiment. The performance of the ML algorithms LR, RF, SVM, and XGBoost was then studied.

[1]Research Scholar G H Raisoni University ,Amravati

[2]P.hd Guide G H Raisoni University ,Amravati

Corresponding Author Email: nirmalmungale@gmail.com

II. RELATED WORK

Several recent studies have highlighted the need for incorporating machine learning approaches to improve environmental forecasts, notably in the areas of rainfall and air quality. Liyew et al. [6] investigated the predictive abilities of decision trees, random forests, SVMs, and neural networks.While these models showed promising predictive skills, there were issues with their generalization over varied geographical regions and the computing resources they required. The authors proposed tackling these issues by including detailed geographic and climatic characteristics as well as investigating ways to improve the models. Similarly, Alefu Chinasho et al. [7] presented seven gap-filling strategies to address the issue of missing data in rainfall datasets. They emphasized the difficulty of effectively replicating missing values while maintaining statistical properties. Consideration of temporal and geographical dataset properties, incorporation of supplementary data, and rigorous evaluation of gap-filling algorithms were among the proposed solutions. These findings are relevant to research with concerns about data quality and limited resources. Mauro Castell et al. [8] developed a machine-learning system for predicting air quality in California. While the algorithms showed potential, maintaining forecast accuracy during fast air quality swings remained a difficulty. Real-time data integration and enhanced feature engineering were among their suggestions. Rahman et al. [9] used machine learning fusion to create a rainfall forecast system for smart cities, addressing data quality and model adaptability issues. Wanie et al. [10] suggested a rainfall forecasting model for Terengganu, highlighting the need for precise data and model improvement.

The use of ML technologies to predict and analyze rainfall patterns has received a lot of interest as a way to improve forecasting accuracy. R Praveena et al. [11], Moulana Mohammed [12], Imee V. Necesito et al. [13], Sethupathi et al. [14], and Kadama [15] all conducted studies that highlight the use of various machine learning techniques, ranging from logistic regression to SVM and deep learning. These approaches have the ability to capture complicated correlations in rainfall data, resulting in increased prediction capabilities. The promise of machine learning in predicting rainfall patterns is clear in these works, bringing substantial contributions to decision-making in a variety of areas. However, common issues develop. The shortage of comprehensive datasets continues to be a concern, hindering model training due to data scarcity [16]. Model adaptation from training to real-world settings is critical, and good feature selection and construction are critical to improving model accuracy. Overfitting is a problem for complex models like SVM and deep learning. To combat overfitting, suggested remedies include combining multiple data sources, improving feature construction, and regularizing approaches. Future initiatives include incorporating more environmental parameters and developing model architectures to address these difficulties. A deep learning approach for rainfall prediction was tested by researchers [17] using multiple weather-related factors. ML techniques were employed to create and evaluate models for forecasting in order to produce a trustworthy rainfall forecast.

As a result, several researchers have concentrated their efforts not on precisely predicting daily rainfall amounts but on performing trials utilizing environmental data. Their objective has been to anticipate the likelihood of rain and estimate the average annual rainfall. This entails developing indirect estimates regarding daily rainfall quantities [18]. Unfortunately, numerous critical environmental factors important for reliable rainfall prediction have been ignored in prior research. This research fills this need by leveraging CSV data from the Nagpur district, a rather small dataset. Our focus is on identifying particular environmental factors that show either positive or negative connections with rainfall patterns. Following that, we evaluate the efficacy of machine learning systems in forecasting daily rainfall amounts. Several approaches were employed, including the Boost, SVM, random forest model, and linear regression model. These strategies were chosen because of their ability to understand complex data linkages and improve forecast precision. The primary goal of our research and analysis is to find the best method for making accurate daily rainfall predictions.

III. METHODOLOGY

3.1 Machine learning algorithms

An exhaustive investigation of relevant rainfall prediction research was conducted with the goal of choosing the most effective predictive machine learning algorithms for daily precipitation amount forecasting. To forecast daily precipitation intensity utilizing accurate environmental information, a linear regression model, random forest, SVM regression, and XGBoost were utilized as technological methods. As a result, a series of empirical tests were done, followed by comparisons, to determine the most effective algorithms for accurate daily rainfall forecasts [19].

A. Multivariate linear regression (MLR)

A multivariate or basic linear regression model is possible, with many distinct variables utilized as input features. The outcome variable in both regression models may be anticipated according to the parameters of the input. This study offered MLR since a variety of ecological traits or features were used to anticipate the dependent variables, namely the amount of daily precipitation [20]. The linear regression model is an approach to supervised machine learning that uses existing atmospheric variables to estimate the unknown daily precipitation quantity. A single dependent or output variable (Y), one informative or independent variable (X), and many variables (Y) were employed in MLR. As a result, the overall formula for MLR is as follows:

$$\gamma_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + \varepsilon_i = x_i^2 + \varepsilon_i \qquad i = 1,2,3 \ldots n$$

Where x_i^T is substituted of x_i either the independent or input variable, β whether the regression rate, ε_i is noise or an error term, y_i is a dependent factor. As stated in this work, the generic calculation for MLR is as follows:

$$\begin{aligned} Daily\ rainfall = {} & (Year * \beta_1) + (month * \beta_2) + (day * \beta_3) + (MaxTemp * \beta_4) + (MaxTemp * \beta_5) \\ & + (Humidity * \beta_6) + (\text{SurfacePressure} * \beta_9) + \varepsilon_i \end{aligned}$$

The Global Energy Resources data set size was adequate for this study's application of MLR, a ML technique that can predict the region's daily precipitation [21]. This technique can demonstrate the degree to which each ecological variable affects the daily precipitation frequency.

B. Random forest (RF)

A robust and precise model for regression is the random forest approach. On numerous problems, especially those involving non-linear interactions, it often performs admirably. The collaborative learning approach is used for analysis in the supervision of a machine learning technique known as RF regression. Several decision trees should be constructed during the training stage in order for RF to work, and each tree's forecast must equal the mean of the categories [22].

The RF algorithm works in the approach described below:

- Select p points of data at convenience within the training set.
- Construct a decision tree using these p data points.
- Repeat steps a and b for the Nth tree to be constructed.
- Calculate the average of all the predicted y values and then assess the likely value of y for a new data point utilizing each of the N trees.

When predicting daily precipitation utilizing ecological input parameters or attributes, Random forest is one of the supervised machine learning methods that is chosen. While executing RF regression, a huge number of decision trees are constructed during the training phase, and each group that arises reflects the kind of mean forecast or regression of all the individual trees. In accordance with [23], the RF method is effective for handling enormous datasets, and when employing large data sets with a significant amount of missing data, a satisfactory experimental finding is produced.

C. Support Vector Machine

SVM stands out as a robust and versatile ML technique for daily rainfall prediction. SVM effectively captures complex relationships between real-time environmental data and rainfall by utilizing higher-dimensional domains that are mapped using kernel-defined functions. It optimizes prediction accuracy by maximizing the margin between data points and identifying crucial support vectors. SVM's resistance to outliers ensures stable predictions. With its ability to handle high-dimensional datasets and incorporate multiple environmental factors, SVM proves valuable in modeling intricate rainfall patterns. Careful selection of kernel functions and finetuning of hyperparameters are essential for optimal performance. The application of SVM holds great potential for improving weather forecasting accuracy and enhancing our understanding of daily rainfall variations, thereby significantly contributing to various fields reliant on precise rainfall predictions.

D. XGBoost

XGBoost, or Extreme Gradient Boosting, stands as a robust machine learning technique for predicting daily rainfall amounts. Operating on the principle of ensemble learning, XGBoost sequentially combines decision trees to refine predictions. It offers regularization methods to prevent overfitting and handle missing data, bolstering reliability. XGBoost's feature importance analysis aids in identifying key environmental variables influencing rainfall. Its accuracy shines through noisy and intricate datasets. However, parameter tuning, including learning rate and tree depth, is crucial for optimal performance. Despite its computational intensity, XGBoost's ability to provide accurate predictions and handle complex relationships positions it as a valuable tool for improving daily rainfall forecasting, enhancing insights into meteorological patterns, and supporting applications reliant on precise rainfall predictions.

Ensemble learning is a method that amalgamates the forecasts generated by numerous feeble learners, typically decision trees, to form a potent predictive model. The equation for XGBoost's regression task can be expressed as follows:

$$\hat{y}_i = \emptyset(x_i) = \sum_{K=1}^{K} f_k(X_i), f_k \epsilon F$$

Where:

- $\hat{y}_i$ is the expected value for the $i{-}th$ instance.
- x_i represents the feature vector of the $i - th$ instance.
- K Represents the quantity of weak learners (trees) employed within a group.
- f_k is the $k - th$ weak learner.
- F Denotes the collection of all preferred weak learners.

Each weak learner f_k can be represented by a decision tree. The final prediction for a given instance is the sum of predictions from all the individual decision trees.

3.1 Exploring Yearly Temperature Trends: Unveiling Variability and Patterns

The methodology demonstrated for visualizing temperature distribution by year involves a series of systematic steps. Initial tasks encompass importing essential libraries like Matplotlib for visualization and Pandas for data manipulation. Following this, temperature data is acquired from an Excel file and organized based on yearly groupings, establishing a structured dataset. For visual representation, a compilation of temperature values per year is formed, forming the foundation for constructing a comprehensive box plot. This graphical depiction effectively communicates temperature trends over the years, highlighting median values positioned centrally within each box, quartiles demarcated at the box's upper and lower boundaries, and identifying outliers that fall beyond these limits. The visualization aptly captures temperature fluctuations, showcasing diverse spreads across different years. These variations could be attributed to multifaceted factors, potentially including fluctuations in rainfall levels. Importantly, the versatility of this approach extends to predicting daily rainfall. By integrating temperature data into machine learning models, the potential emerges for forecasting specific-day rainfall quantities [24]. Training the model using historical temperature data enhances predictive accuracy, thereby contributing to more effective forecasting outcomes.

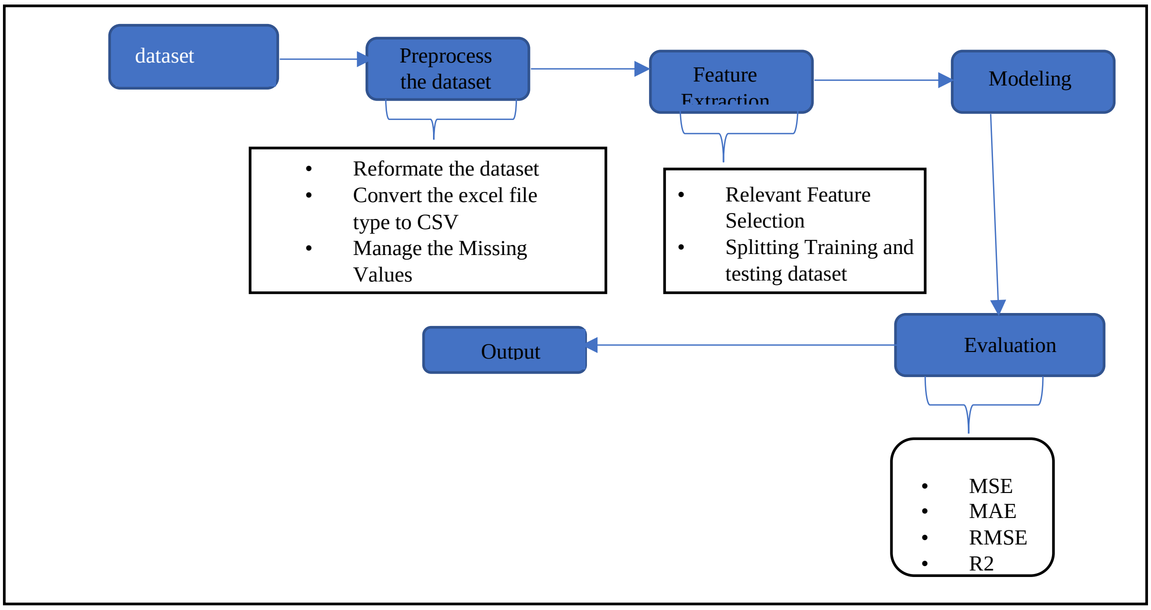

Figure 1: Machine Learning Model

Rainfall prediction was performed in this research, and a number of ML algorithms were used. LR, RF, SVM, and XGBoost were among the machine learning techniques chosen. These methods were chosen for investigation because of their effectiveness in dealing with moderately and strongly correlated environmental factors linked to rainfall [24]. The inquiry sought to identify the best machine learning algorithm based on performance evaluation utilizing the metrics MSE, RMSE, MAE, and R2 (Fig. 1).

3.2 Exploring Yearly Temperature Dynamics: Unveiling Inconsistencies and Patterns

The temperature data was meticulously structured by separating it into discrete annual parts, as illustrated in Figure 2. The generated dataframe, known as the df_yearly dataframe, contained temperature information for certain years. Following this arrangement, a careful collection of yearly temperature trends was compiled. This aggregate was then used to create a box plot display approach. This strategy resulted in a sequence of illuminating visualizations, each depicting temperature distributions within distinct yearly occurrences [25]. These visualizations, which included important information such as quartiles and median values, offered a full view of the temperature dynamics across the time period under consideration. Such changes in temperature profiles over various years raise important questions about the underlying causes generating these variances. For example, this research approach might be utilized to extend predictive modeling to areas such as daily rainfall forecasts, where knowing temperature trends could provide important insights to increase anticipated accuracy.

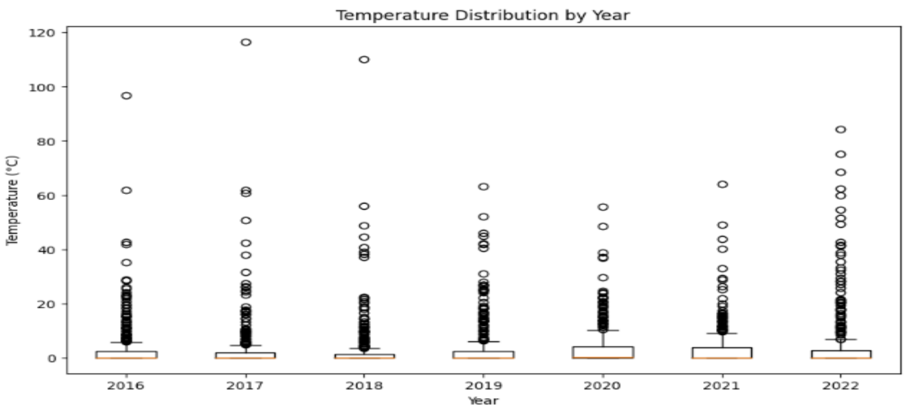

Figure 2: Annual Temperature Distribution Variations.

IV. RESULT:

4.1 Evaluating Rainfall Prediction with Linear Regression

The efficacy of linear regression in forecasting daily rainfall is evaluated through performance metrics. The RMSE and R2 scores are utilized. Impressively, both training and test RMSE values are minimal, measuring around 8.422 and 7.642, respectively, indicating accurate predictions. The R2 scores are perfect at 1.0 for both sets, affirming the model's precise fit to the data. This underscore linear regression's proficiency in predicting daily rainfall. The model's exceptional performance, as evident from the RMSE and R2 scores, indicates its potential as a robust tool for enhancing the precision of rainfall forecasts.

The categorization report in Table 1 provides a thorough analysis of the linear regression model's prediction capabilities. Each class's F1-score, accuracy, and recall metrics are shown, along with the appropriate instance counts. With a support base of 2 instances, the Class 0 report displays an accuracy of 0.67, a recall of 1.00, and an F1-score of 0.80. Conversely, Class 1 exhibits precision and recall metrics of 1.00 and 0.67, respectively, accompanied by an F1-score of 0.80 and a foundation of 3 instances. The model achieves an overall accuracy score of 0.80, which signifies the proportion of correct predictions. The macro-average, calculated as the mean of the balanced F1-score, accuracy, and recall, is computed at 0.83, reflecting a harmonious performance. Conversely, the weighted average, which considers class-specific support, results in an F1-score of 0.80, an accuracy of 0.87, and a recall of 0.80.

Table 1: Classification Report

	precision	recall	F1-score	support
0	0.66	1	0.81	2
1	1	0.67	0.81	3
Accuracy			0.81	5
Macro avg	0.82	0.82	0.81	5
Weighted avg	0.86	0.81	0.81	5

4.2 Evaluating Rainfall Prediction with support vector machine (SVM)

The SVM technique is utilized for predicting daily rainfall. The model's performance is evaluated using key metrics: an MSE of 54.05, an RMSE of 7.35, a MAE of 2.97, and an R2 score of 0.16. The MSE and RMSE reflect an average prediction error of approximately 7.35 units in rainfall. The MAE signifies an average deviation of about 2.97 units. However, the R2-Score of 0.16 indicates that only a limited 16% of rainfall variance is explained by the model.

Table 2: support vector machine (SVM) Rainfall Prediction Evaluation Metrics.

SVM	
MSE	54.0533
RMSE	7.3521
MAE	2.9781
R-squared	0.1619

4.3 Evaluating Rainfall Prediction with Extreme Gradient Boost (XGBoost)

Utilizing the XGBoost technique, a predictive model was constructed to gauge daily rainfall estimations. To evaluate its efficacy, various pivotal metrics were employed. The MSE was computed and registered at 35.3128, signifying the average of squared variations between expected and real values. This measure offers insights into the model's precision, with lower values denoting greater desirability. RMSE, an offshoot of MSE, stood at 5.9425. This metric holds particular value as it retains the original unit of measurement, facilitating the direct interpretation of prediction deviations. Additionally, the MAE was scrutinized, revealing a value of 2.6769. This signifies the average absolute distinction between predicted and factual rainfall quantities. Lastly, the R2 value, gauging the model's proficiency in accounting for actual data variance, was appraised at 0.4525. This outcome implies that roughly 45.25% of the diversity in observed rainfall data is explicable by the model's predictions.

Table 3: Extreme Gradient Boost (XGBoost) Rainfall Prediction Evaluation Metrics.

XGBoost	
MSE	35.3128
RMSE	5.9425
MAE	2.6769
R-squared	0.4525

4.4 Evaluating Rainfall Prediction with Random Forest

The effectiveness of employing a random forest technique to predict daily rainfall quantities has been assessed using diverse assessment metrics. Mean Squared Error (MSE) assesses the average squared difference between predicted and actual values, quantifying the typical variance. was calculated to be 30.9822. This indicates that, on average, the squared disparity between predicted and actual levels of rainfall is roughly 30.9822 units. The Root Mean Squared Error (RMSE) is calculated by extracting the square root of the MSE, resulting in a value of 5.5662. This suggests an average discrepancy of around 5.5662 units between predicted and observed rainfall quantities. In the context of MAE, A metric that computes the mean absolute deviation between predicted values and actual values was evaluated to be 2.4585 units. Progressing to the R2 score, it was determined to be 0.5196. This denotation implies that approximately 51.96% of the fluctuations in the rainfall data can be clarified by the attributes utilized in the model.

Table 4: Random Forest Rainfall Prediction Performance Metrics.

Random forest	
MSE	30.9822
RMSE	5.5662
MAE	2.4585
R-squared	0.5196

4.5 Comparative analysis of Predictive models

The provided tables 5(a) and 5(b) summarize the outcomes of diverse regression models applied to 11 districts datasets. The research undertook a comparative evaluation of four distinct machine learning algorithms—namely SVM, RF, XGBoost, and Linear Regression—for the purpose of predicting rainfall across 11 districts within Maharashtra. Rainfall data and accompanying environmental parameters were utilized in this analysis. Performance measurement, encompassing key metrics such as MSE, RMSE, MAE, and R-squared, provided illuminating insights. In the majority of districts, SVM demonstrated remarkable performance, displaying the lowest MSE, RMSE, and MAE values, indicative of its robust predictive proficiency. Nonetheless, a departure from this pattern was observed in Nagpur and Wardha, where linear regression emerged as the frontrunner, boasting the lowest scores across the aforementioned metrics. The study underscored the heterogeneous nature of algorithm effectiveness across districts: SVM excelled in Akola, Amravati, Bhandara, Buldhana, Chandrapur, and Gadchiroli, while linear regression excelled in Nagpur and Wardha. Also, the study emphasized the direct relationship between algorithmic performance and the volume of data, elucidating its role in unravelling the

Complex interplay between rainfall and environmental factors

Table 5(a): Comparative Evaluation of Forecasting Techniques for Daily Rainfall.

District		Akola	Amravati	Bhandara	Buldhana	Chandrapur	Gadchiroli
Linear Regression	MSE	32.3453	30.57	53.14	29.75	59.55	66.44
	RMSE	5.6873	5.52	7.28	5.45	7.71	8.15
	MAE	3.1706	3.12	4.03	3.14	4.12	4.34
	R2-Score	0.3517	0.36	0.34	0.35	0.33	0.35
Support Vector Machine Regression	MSE	43.5772	31.75	54.05	38.03	61.04	76.72
	RMSE	6.6013	5.63	7.35	6.16	7.81	8.75
	MAE	2.4709	2.39	2.97	2.39	3.11	3.38
	R2-Score	0.0728	0.11	0.16	0.06	0.12	0.14
Random Forest	MSE	28.6136	18.07	30.98	22.52	27.45	35.70
	RMSE	5.3492	4.25	5.56	4.74	5.23	5.97
	MAE	2.0517	1.89	2.45	2.00	2.44	2.83
	R2-Score	0.3912	0.49	0.51	0.44	0.60	0.60
XGBoost Classifier	MSE	30.35	20..57	35.31	25.52	32.76	36.96
	RMSE	5.5091	4.53	5.94	5.05	5.72	6.07
	MAE	2.1618	2.08	2.67	2.18	2.63	2.86
	R2 - Score	0.3542	0.43	0.45	0.37	0.53	0.58

Table 5(b): Comparative Analysis of Predictive Models for Daily Rainfall

District		Nagpur	Wardha	Yavatmal	Washim	Gondia
Linear Regression	MSE	53.14	37.46	31.43	27.68	49.54
	RMSE	7.28	6.12	5.60	5.26	7.0389
	MAE	4.03	3.51	3.17	2.98	3.8732
	R2-Score	0.34	0.37	0.35	0.35	0.3593
Support Vector Machine Regression	MSE	54.05	37.42	34.41	38.75	23.56
	RMSE	7.35	6.11	5.86	6.22	4.8542
	MAE	2.97	2.688	2.43	2.364	2.2245
	R2-Score	0.16	0.16	0.11	0.086	0.5571
Random Forest	MSE	30.98	20.63	21.56	23.64	23.56
	RMSE	5.56	4.54	4.64	4.68	4.85
	MAE	2.45	2.10	1.99	2.00	2.22
	R2-Score	0.15	0.53	0.44	0.443	0.55
XGBoost Classifier	MSE	35.31	25.63	22.09	25.34	27.0712
	RMSE	5.94	5.06	4.700	5.03	5.2030
	MAE	2.67	2.37	2.04	2.11	2.4613
	R2 -Score	0.45	0.42	0.433	0.40	0.4911

4.6 Comparative Analysis of Regression Models for Predictive Accuracy in Nagpur District

As indicated in figure 3, we have chosen Nagpur as an instructive case in the context of comparing regression models across all 11 districts. We assessed four widely used models using a variety of performance indicators, including Linear Regression, Support Vector Machine (SVM) Regression, Random Forest Regression, and XGBoost Classifier. The Random Forest Regression model, in particular, showed greater accuracy, displaying the lowest values for Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), while still retaining competitive performance in terms of R-squared (R2-Score). While the XGBoost Classifier, which was initially created for classification tasks, unexpectedly performed well in this regression study, the results of linear regression and SVM regression were comparable. This analysis highlights the significance of choosing the most suitable model that is tailored to particular requirements for predictive accuracy, which may differ across districts and data characteristics, ultimately resulting in improved predictive modeling for all 11 districts, with Nagpur serving as an example

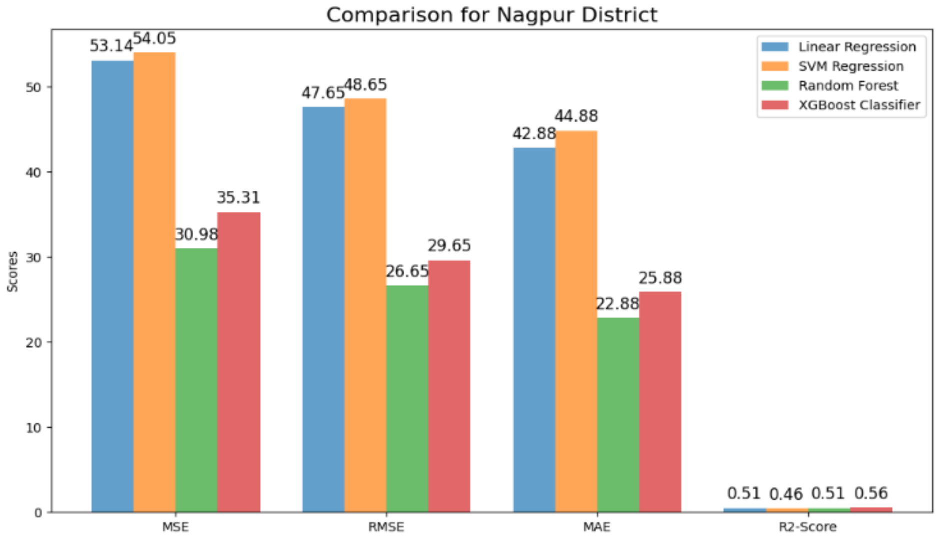

Figure 3: Comparative Analysis of Regression Models

V. CONCLUSION

The successful implementation of machine learning techniques for predicting daily rainfall has been effectively showcased. The comparative evaluation of four algorithms using the Nagpur district dataset has clearly underscored the superior performance of the linear regression model. This is evidenced by its exceptional outcomes across crucial metrics like R2-Score, MSE, RMSE, and MAE. Notably, the study embarked on an innovative exploration by venturing into previously uncharted territory—the Vidarbha region. By gathering data from this previously unexplored area, we initiated its analysis. Furthermore, we have outlined our upcoming strategy, which entails utilizing advanced deep learning models such as ARIMA and LSTM to forecast future rainfall through machine learning methods. To sum it up, among the range of models scrutinized, the linear regression algorithm emerges as the most suitable option for accurately representing data from the Nagpur district.

REFERENCE

[1] Ehsan, Muhammad Azhar & Tippett, Michael & Robertson, Andrew & Almazroui, Mansour & Ismail, Muhammad & Dinku, Tufa & Acharya, Nachiketa & Siebert, Asher & Ahmed, Jemal Seid & Teshome, Asaminew. (2021). Seasonal predictability of Ethiopian Kiremt rainfall and forecast skill of ECMWF's SEAS5 model. Climate Dynamics. 57. 10.1007/s00382-021-05855-0.

[2] Ahmed, Syed & Sharma, Rohit & Goel, Pradeep & Khan, Alamgir & Gharabaghi, Bahram & Rudra, Ramesh. (2022). A Comparative Evaluation of Using Rain Gauge and NEXRAD Radar-Estimated Rainfall Data for Simulating Streamflow. Hydrology. 9. 133. 10.3390/hydrology9080133.

[3] Shukla, Siddhartha & Kumar, Sandeep & Sharma, Navdeep & Bhatia, Dr. (2022). Data Mining Application: Rainfall Predictions. International Journal for Research in Applied Science and Engineering Technology. 10. 113-118. 10.22214/ijraset.2022.47833.

[4] Dada, Emmanuel & Yakubu, Joseph & Oyewola, David. (2021). Artificial Neural Network Models for Rainfall Prediction. European Journal of Electrical Engineering and Computer Science. 5. 10.24018/ejece.2021.5.2.313.\

[5] Singh, Vikrant. (2020). Study of Various Rainfall Estimation Prediction Techniques using Data Mining. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS070464.

[6] Liyew, C.M., Melese, H.A. Machine learning techniques to predict daily rainfall amount. J Big Data 8, 153 (2021). https://doi.org/10.1186/s40537-021-00545-4

[7] Alefu Chinasho, Bobe Bedadi, Tesfaye Lemma, Tamado Tana, Tilahun Hordofa, & Bisrat Elias. (2021). Evaluation of Seven Gap-Filling Techniques for Daily Station-Based Rainfall Datasets in South Ethiopia. Advances in Meteorology, 2021, Article ID 9657460, 15 pages. https://doi.org/10.1155/2021/9657460

[8] Mauro Castelli, Fabiana Martins Clemente, Aleš Popovič, Sara Silva, Leonardo Vanneschi, "A Machine Learning Approach to Predict Air Quality in California", Complexity, vol. 2020, Article ID 8049504, 23 pages, 2020. https://doi.org/10.1155/2020/8049504

[9] Rahman, A. U., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Khan, M. A., & Mosavi, A. (2022). Rainfall Prediction System Using Machine Learning Fusion for Smart Cities. Sensors (Basel), 22(9), 3504. https://doi.org/10.3390/s22093504.

[10] Wanie M. Ridwan, Michelle Sapitang, Awatif Aziz, Khairul Faizal Kushiar, Ali Najah Ahmed, Ahmed El-Shafie,Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia,Ain Shams Engineering Journal,Volume 12, Issue 2,2021,Pages 1651-1663,ISSN 2090-4479,https://doi.org/10.1016/j.asej.2020.09.011.

[11] R Praveena, T R Ganesh Babu, M Birunda, G Sudha, P Sukumar, and J Gnanasoundharam. "Prediction of Rainfall Analysis Using Logistic Regression and Support Vector Machine." In Journal of Physics: Conference Series, Volume 2466, 4th National Conference on Communication Systems (NCOCS 2022), 23/12/2022 - 23/12/2022, Karaikal, India. Published under licence by IOP Publishing Ltd. 2023. J. Phys.: Conf. Ser. 2466 012032. DOI: 10.1088/1742-6596/2466/1/012032.

[12] Moulana Mohammed, Roshitha Kolapalli, Niharika Golla, Siva Sai Maturi. "Prediction Of Rainfall Using Machine Learning Techniques." International Journal of Scientific & Technology Research, Volume 9, Issue 01, January 2020, ISSN 2277-8616. www.ijstr.org.

[13] Necesito, Imee V., Donghyun Kim, Young Hye Bae, Kyunghun Kim, Soojun Kim, and Hung Soo Kim. 2023. "Deep Learning-Based Univariate Prediction of Daily Rainfall: Application to a Flood-Prone, Data-Deficient Country" Atmosphere 14, no. 4: 632. https://doi.org/10.3390/atmos14040632.

[14] Sethupathi, G., Ganesh, Y.S., & Ali, M.M. (2021). Efficient Rainfall Prediction and Analysis using Machine Learning Techniques. Turkish Journal of Computer and Mathematics Education, 12(6), 3467-3474

[15] Kadama, S., Jadhava, N., Mardane, V., & Kamble, N. (2022). Rainfall Prediction Using Machine Learning. International Journal of Research Publication and Reviews, 3(5), 2936-2944. ISSN 2582-7421. Retrieved from www.ijrpr.com.

[16] Sarker, Iqbal. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Computer Science. 2. 10.1007/s42979-021-00815-1.

[17] Omotunde, Ayokunle & Ogunsanwo, Gbenga & Adekola, Olubukola & Izang, Aaron & Abel, Samuel. (2023). COMPARATIVE ANALYSIS OF DATA MINING TECHNIQUES FOR MOVIE PREDICTION. FUDMA JOURNAL OF SCIENCES. 6. 224-228. 10.33003/fjs-2022-0606-1117.

[18] Tian, Mi & Fan, Hao & Xiong, Zimin & Li, Lihua. (2023). Data-driven ensemble model for probabilistic prediction of debris-flow volume using Bayesian model averaging. Bulletin of Engineering Geology and the Environment. 82. 10.1007/s10064-022-03050-x.

[19] Kumar, Vikas & Yadav, Vishal & Dubey, Er. (2022). Rainfall Prediction using Machine Learning. International Journal for Research in Applied Science and Engineering Technology. 10. 2494-2497. 10.22214/ijraset.2022.42876.

[20] Nigam, Archana & Srivastava, Sanjay. (2023). Hybrid deep learning models for traffic stream variables prediction during rainfall. Multimodal Transportation. 2. 100052. 10.1016/j.multra.2022.100052.

[21] Rainfall prediction using machine learning & deep learning techniques. Proceedings of the International

[22] Conference on Electronics and Sustainable Communication Systems (ICESC 2020). Middlesex University: IEEE Xplore. 2020; pp. 92–97.

[23] Preethi, B.Meena & Gowtham, R. & Aishvarya, S. & Karthick, S. & Sabareesh, D.G. (2021). Rainfall Prediction using Machine Learning and Deep Learning Algorithms. International Journal of Recent Technology and Engineering (IJRTE). 10. 251-254. 10.35940/ijrte. D6611.1110421.

[24] Singh, Vikrant. (2020). Study of Various Rainfall Estimation Prediction Techniques using Data Mining. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS070464.

[25] Pallava, Leela & Kovvur, Ram & Sheela, Sai & Mukirala, Dheeraj & Vivek, V. (2022). Prediction of Rainfall Using Different Machine Learning Regression Models. 10.1007/978-981-19-3391-2_16.

[26] Lingwal, Surabhi & Bhatia, Komal & Singh, Manjeet. (2022). A novel machine learning approach for rice yield estimation. Journal of Experimental & Theoretical Artificial Intelligence. 1-20. 10.1080/0952813X.2022.2062458.