International Journal of Advanced Technology in Engineering and Science
Vol. No. 12, Issue No. 01, January 2024
www.ijates.com
ijates
ISSN 2348 - 7550

Precise Rainfall Prediction Using Ensemble Learning

[1]Vijaya Lakshmi D M

Assistant Professor/CSE

Adhiyamaan College of Engineering (Autonomous), Hosur.

[2]Bharathwaj V

UG Student,

Adhiyamaan College of Engineering (Autonomous), Hosur.

[3]Gokulnath R

UG Student

Adhiyamaan College of Engineering (Autonomous), Hosur.

[4]Jacobraj A

UG Student

Adhiyamaan College of Engineering (Autonomous), Hosur.

Abstract- Precise Rainfall Prediction (in mm) Using Ensemble Learning introduces an ensemble model that integrates the predictive capabilities of Ada Boost Regressor, Gradient Boosting Regressor, and Random Forest Regressor. Using the distinctive strengths of these individual models, the model enhances the accuracy of rainfall predictions in the 115-year dataset (1901-2015) for India. Through hyperparameter tuning and model evaluation, our ensemble learning system demonstrates its efficiency in forecasting rainfall prediction across india for the next 5 years. This solution contributes the government of india to compare the data with the existing models to improve decision-making across diverse sectors such as Waste Water Treatment Plan (WWTP), Hydroelectric energy reservoirs etc. that are reliant on precise and timely precipitation information.

Keywords - Rainfall prediction, Ensemble learning, Ada Boost Regressor, Gradient Boosting Regressor, Random Forest Regressor, hyperparameter tuning, model evaluation, Waste Water Treatment Plan (WWTP), Hydroelectric energy.

Introduction

Rainfall prediction is a critical facet of climate science with profound implications for various sectors, ranging from agriculture to energy management. As the world grapples with the challenges posed by climate variability, the demand for accurate and timely precipitation forecasts has intensified. In this context, our research addresses the pivotal task of predicting rainfall in India over the next five years, utilizing an ensemble learning approach. Leveraging the distinct strengths of Ada Boost Regressor, Gradient Boosting Regressor, Random Forest Regressor, and Voting Regressor, our ensemble model aims to surpass the limitations of individual algorithms. The rich dataset spanning 115 years (1901-2015) provides a robust foundation for training and evaluating our predictive system. Through meticulous hyperparameter tuning and comprehensive model evaluation, our study not only contributes to advancing the precision of rainfall forecasts but also presents a practical solution for the Government of India to enhance decision-making across sectors dependent on precise meteorological insights. This paper unfolds the methodology, results, and implications of our ensemble learning system, underscoring its potential impact on crucial sectors such as Waste Water Treatment Plants and Hydroelectric energy reservoirs.

Related Works

This review explores the application of ensemble learning algorithms, such as bagging, boosting, and stacking, in the prediction of rainfall. Given the significant impact of heavy precipitation on human life and agriculture, accurate rainfall forecasts are crucial. The study reveals that ensemble techniques outperform individual models, with boosting methods like AdaBoost and extreme gradient boosting being particularly effective in rainfall prediction scenarios [1]. This paper addresses the challenges of non-stationarity in rainfall time series for precise forecasting in the face of climate change. The proposed hybrid CEEMD-RF-KRR model, integrating complete ensemble empirical mode decomposition with Random Forest and Kernel Ridge Regression algorithms, demonstrates superior performance in rainfall prediction at multiple locations in Pakistan. The model achieves high accuracy, making it valuable for agriculture, water resource management, and early warning systems for droughts and floods [2]. This research demonstrates the superior predictive performance of machine learning algorithms, including Genetic Programming, Support Vector Regression, and Neural Networks, over the current state-of-the-art techniques, such as Markov chain extended with rainfall prediction, in the challenging task of rainfall prediction. The study encompasses 42 cities with diverse climatic features, revealing the ability of machine learning methods to outperform existing approaches and detect correlations between climates and predictive accuracy. Overall, this work highlights the significant positive impact of machine learning-based intelligent systems on rainfall prediction [3]. This study explores the effectiveness of random forests ensemble classification and regression in improving hourly rainfall rate assignment using Meteosat Second Generation (MSG) SEVIRI data. The model incorporates cloud physical properties, including water vapor-IR differences, IR cloud top temperature, spectral SEVIRI channels, and cloud properties. The developed technique demonstrates good accuracy in predicting rainfall rates on an hourly basis during day, night, and twilight conditions, offering a comprehensive 24-hour precipitation estimation approach [4]. This study explores the effectiveness of a hybrid forecasting model, RSVR (combining Recurrent Neural Networks and Support Vector Machines), to accurately predict rainfall depth values. The model utilizes the Chaotic Particle Swarm Optimization algorithm for parameter selection in the Support Vector Regression (SVR) component. The empirical results, based on rainfall data from Northern Taiwan during typhoon periods, demonstrate the superior forecasting performance of the proposed RSVRCPSO model, suggesting it as a promising alternative for rainfall prediction [5].

Existing System

The existing system for rainfall prediction primarily relies on conventional meteorological models and statistical techniques. These traditional approaches often face challenges in accurately capturing the complex and nonlinear patterns inherent in rainfall data. Classical regression models and time-series analyses may struggle to account for the intricate dependencies present in historical rainfall datasets spanning over a century. Furthermore, these methods may not effectively adapt to changing climatic conditions and regional variations, limiting their accuracy and predictive capabilities. The reliance on single models, without the integration of diverse approaches, can hinder the ability to grasp the nuances of precipitation patterns across different geographical regions. Overall, the limitations of existing systems underscore the need for more sophisticated and adaptable methodologies to enhance the precision of rainfall predictions, especially considering the critical implications of such forecasts for various sectors like agriculture, water resource management, and energy production.

Proposed System

The proposed system aims to revolutionize rainfall prediction by introducing an ensemble model that seamlessly integrates the predictive capabilities of AdaBoostRegressor, GradientBoostingRegressor, RandomForestRegressor, and VotingRegressor. This comprehensive approach leverages the distinctive strengths of each individual model to create a robust and accurate forecasting system. The system utilizes a 115-year dataset spanning from 1901 to 2015, providing a rich historical context for predicting rainfall patterns in India. To optimize the ensemble's performance, extensive hyperparameter tuning is employed, ensuring the models are finely tuned for the unique characteristics of the dataset. Rigorous model evaluation processes validate the system's efficacy, establishing its reliability in forecasting rainfall for the next 5 years. By contributing precise and timely precipitation information, this proposed system not only advances the field of rainfall prediction but also holds the potential to significantly impact decision-making processes across diverse sectors in India, including Waste Water Treatment Plants and Hydroelectric energy reservoirs.

Methodology

Data Collection

In this study, we sourced historical rainfall data for India spanning 115 years, from 1901 to 2015. The dataset comprises monthly and seasonal rainfall measurements recorded in millimeters. The data was obtained from reliable meteorological sources and underwent a thorough validation process to ensure accuracy and consistency. Special attention was given to the geographical diversity of the data, incorporating information from various regions across India to ensure a representative and comprehensive dataset for training and testing our models.

Dataset

REGION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
INDIA	1901	34.7	37.7	18	39.3	50.8	113.4	242.2	272.9	124.4	52.7	38	8.3	1032.3	72.4	108.1	752.8	99
INDIA	1902	7.4	4.3	19	43.5	48.3	108.8	284	199.7	201.5	61.5	27.9	24.4	1030.2	11.7	110.8	794	113.8
INDIA	1903	17	8.3	31.3	17.1	59.5	118.3	297	270.4	199.1	117.9	36.9	17.7	1190.5	25.3	107.9	884.8	172.5
INDIA	1904	14.4	9.6	31.8	33.1	72.4	164.8	261	206.4	129.6	69	11.2	16.3	1019.8	24	137.4	761.8	96.6
INDIA	1905	25.3	20.9	42.7	33.7	55.7	93.3	252.8	200.8	178.4	51.4	9.7	10.5	975.3	46.2	132.2	725.4	71.6
INDIA	1906	21.2	50.8	31.7	16.1	36.7	177.1	285.4	251.1	182.7	49.6	17.8	23.9	1144.1	72	84.5	896.3	91.3
INDIA	1907	16.2	46	37.8	62.8	32.6	154.4	225.4	310.4	96.9	22.7	22.5	12.1	1039.7	62.1	133.1	787.1	57.3
INDIA	1908	21.8	17.8	8.7	33.3	45.4	126.4	327.2	312.8	150.5	37.5	6.5	8.7	1096.6	39.6	87.4	916.9	52.7
INDIA	1909	23.3	17.9	7.2	63	52.1	208.2	309.4	235	163.6	38.3	9.8	30.2	1158.1	41.3	122.3	916.2	78.3
INDIA	1910	15.6	11.1	13.8	29.2	41.5	213.5	250.3	289.8	186.8	107.4	35	6.4	1200.3	26.7	84.5	940.3	148.8
INDIA	1911	45.7	5.6	49.9	22.8	47.6	191.9	162.7	213.5	182.3	70.6	42.8	12	1047.5	51.3	120.3	750.4	125.5
INDIA	1912	22.1	21.8	21.2	40	43	106.8	329	261.3	120.3	55.9	50.8	5.8	1078.1	44	104.1	817.5	112.5
INDIA	1913	6.6	40.8	25	26.1	73.2	214.4	272.4	193.9	108.4	67	17.4	23.9	1068.9	47.4	124.3	789	108.2
INDIA	1914	4.8	31.5	28.3	46.3	69.5	159.3	347.3	240.8	194.3	48.5	21.1	22.5	1214.3	36.4	144.1	941.7	92.2
INDIA	1915	19.4	41.6	46.2	36.4	62.7	155.9	230	230.1	172	90.3	44.1	8.5	1137.2	61	145.3	788	142.9
INDIA	1916	5.1	22.7	12.3	36.1	59.5	235.2	267.1	310.5	201.6	138.2	45.6	3.1	1337.1	27.9	108	1014.4	186.9
INDIA	1917	8.7	38.7	22.8	43.2	75	231.8	285.2	296.5	281	158.8	28.2	10.3	1480.3	47.3	141.1	1094.5	197.3
INDIA	1918	12.2	4.4	41.6	38.8	102.8	212.6	183.8	242.7	109.7	20	41.1	16.4	1026.2	16.5	183.2	748.8	77.6
INDIA	1919	52.5	22.7	20.3	33.5	60.5	195.2	309.3	289.5	163.6	92.5	49.8	19.5	1308.9	75.2	114.3	957.6	161.8
INDIA	1920	25.1	23.9	58.2	38.8	53.9	163.9	297.5	191.7	123	44.1	24.5	3.3	1047.9	49	150.9	776	71.9
INDIA	1921	40.6	8.8	18.6	43.5	49.4	193.9	297.5	278.5	206.6	71.1	14.5	17.2	1240.2	49.4	111.5	976.5	102.8
INDIA	1922	29.7	10.3	14.6	31.5	49	202.7	317.8	222.7	202.7	61	55.5	14.5	1211.9	39.9	95.1	945.9	130.9
INDIA	1923	24.5	41.5	22.7	32.1	60.7	102.7	339.5	277	178.7	60.9	19.4	16.6	1176.3	66	115.5	897.9	97
INDIA	1924	23.1	25.5	14.3	31.7	64.4	135.5	328	263.1	242.5	66.2	57.9	16.1	1268.3	48.6	110.4	969.1	140.2
INDIA	1925	14.2	12.5	15.8	44	104.1	209.2	306.8	238.9	141.7	68.8	39.9	14.4	1210.4	26.7	163.9	896.7	123.1
INDIA	1926	29.2	11.5	60.1	42.7	58.6	99	320.9	335.5	211	57.3	10.6	10.4	1246.8	40.7	161.4	966.4	78.3
INDIA	1927	14.6	37	23.1	37.4	51.8	179.4	351	259.3	173.1	69.2	55.9	10.8	1262.6	51.6	112.3	962.8	135.9
INDIA	1928	22.9	41.9	22.7	34.6	54.3	179.1	306.6	232.8	144.4	128.8	23.6	26.6	1218.4	64.8	111.7	862.9	179

Data Preprocessing

Prior to model training, a comprehensive data preprocessing pipeline was implemented. This included handling missing values, outlier detection, and normalization of features. Seasonal data, divided into Jan-Feb, Mar-May,

ijates
ISSN 2348 - 7550

Jun-Sep, and Oct-Dec, was appropriately formatted to capture the temporal patterns effectively. To enhance the model's ability to generalize, categorical variables were encoded, and temporal features were carefully engineered. Additionally, the dataset was split into training and testing sets to facilitate robust model evaluation.

Handling Missing Values

Identify and fill in or remove any missing values in your dataset to avoid issues during model training.

Outlier Detection

Identify and handle extreme values (outliers) that could skew the analysis or model predictions.

Normalization

Scale numerical features to a standard range, typically between 0 and 1, to ensure that all variables have equal influence on the model.

Encoding Categorical Variables

Convert categorical variables (like 'Yes'/'No' or 'Red'/'Blue'/'Green') into numerical format for the model to understand.

Feature Engineering

Create new features or modify existing ones to better represent the patterns in the data and improve the model's performance.

Train-Test Split

Divide your dataset into two parts: one for training the model and one for testing its performance. This helps ensure the model generalizes well to new, unseen data.

Ensemble Learning Models

The ensemble model utilized in this study integrates three powerful regression algorithms: AdaBoostRegressor, GradientBoostingRegressor, and RandomForestRegressor, along with a VotingRegressor for model aggregation. Each individual model was fine-tuned through an extensive hyperparameter search. AdaBoostRegressor's emphasis on weak learners, GradientBoostingRegressor's boosting approach, and RandomForestRegressor's ensemble of decision trees contribute unique strengths to the combined model. The VotingRegressor facilitates the aggregation of predictions, leveraging the diverse expertise of each base model for superior overall performance.

```
import pandas as pd
from sklearn.ensemble import AdaBoostRegressor, GradientBoostingRegressor, RandomForestRegressor, VotingRegressor
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import mean_absolute_error, r2_score
import matplotlib.pyplot as plt

# Load your dataset (replace 'your_dataset.csv' with the actual file path)
df = pd.read_csv("rainfall_data.csv")

# Assuming 'REGION' is not a feature for prediction
features = ['YEAR', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV', 'DEC', 'Jan-Feb',

# Split the data into training and testing sets
```

Model Training and Validation

The training process involved fitting each base model to the training dataset and optimizing their respective hyperparameters. Cross-validation techniques, such as k-fold cross-validation, were employed to assess model generalization and mitigate overfitting. Model performance metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared, were carefully monitored during training to guide the fine-tuning process. The final ensemble model was then trained on the entire dataset and evaluated on the reserved test set to ensure its predictive capabilities on unseen data.

```
regressor1 = AdaBoostRegressor(random_state=42)
regressor2 = GradientBoostingRegressor(random_state=42)
regressor3 = RandomForestRegressor(random_state=42)

# Define hyperparameter grids for each regressor
param_grid1 = {'n_estimators': [50, 100, 150], 'learning_rate': [0.01, 0.1, 0.2]}
param_grid2 = {'n_estimators': [50, 100, 150], 'learning_rate': [0.01, 0.1, 0.2]}
param_grid3 = {'n_estimators': [50, 100, 150], 'max_depth': [None, 10, 20], 'min_samples_split': [2, 5, 10]}

# Create GridSearchCV for each regressor
grid_search1 = GridSearchCV(regressor1, param_grid1, cv=5, scoring='neg_mean_absolute_error', n_jobs=-1)
grid_search2 = GridSearchCV(regressor2, param_grid2, cv=5, scoring='neg_mean_absolute_error', n_jobs=-1)
grid_search3 = GridSearchCV(regressor3, param_grid3, cv=5, scoring='neg_mean_absolute_error', n_jobs=-1)

# Fit each model with the parameter grid
grid_search1.fit(X_train, y_train)
grid_search2.fit(X_train, y_train)
grid_search3.fit(X_train, y_train)

# Get the best parameters for each regressor
best_params1 = grid_search1.best_params_
best_params2 = grid_search2.best_params_
best_params3 = grid_search3.best_params_

# Print the best parameters for each regressor
print("Best Hyperparameters for AdaBoostRegressor:", best_params1)
print("Best Hyperparameters for GradientBoostingRegressor:", best_params2)
print("Best Hyperparameters for RandomForestRegressor:", best_params3)
```

Model Evaluation and Comparison

To assess the performance of the ensemble model, a thorough evaluation was conducted, comparing its predictions against those of individual models and benchmarking against existing models. Statistical significance tests were applied to validate the improvement achieved by the ensemble approach. Moreover, interpretability analyses were performed to gain insights into the contributions of each base model. The evaluation process aimed to provide a comprehensive understanding of the ensemble model's strengths and limitations in the context of rainfall prediction.

```
# Print actual vs predicted rainfall for each month in 2016
print("Actual vs Predicted Rainfall for Each Month in 2023:")
for i in range(1, 14):
    month_name = df.columns[i]
    actual_value = y_test.iloc[i-1]
    predicted_value = predictions_2016[i-1]
    print(f"{month_name}: Actual - {actual_value} mm, Predicted - {predicted_value} mm")

# Evaluate the model for the entire test set
mae = mean_absolute_error(y_test, predictions_2016)
r2 = r2_score(y_test, predictions_2016)

# Print evaluation metrics for the entire test set
print("\nEvaluation Metrics for the Entire Test Set:")
print(f"Mean Absolute Error (MAE): {mae}")
print(f"R-squared Score: {r2}")
```

LSTM Model

In the context of our rainfall prediction project, the development of the Long Short-Term Memory (LSTM) model involved harnessing the power of recurrent neural networks (RNNs) to capture intricate temporal dependencies within the historical rainfall dataset.

This sophisticated model architecture excels in handling sequences of data, making it particularly well-suited for time series forecasting. The LSTM model was constructed with multiple memory cells capable of retaining and selectively discarding information over extended periods, allowing it to capture long-range dependencies in the monthly and seasonal rainfall patterns.

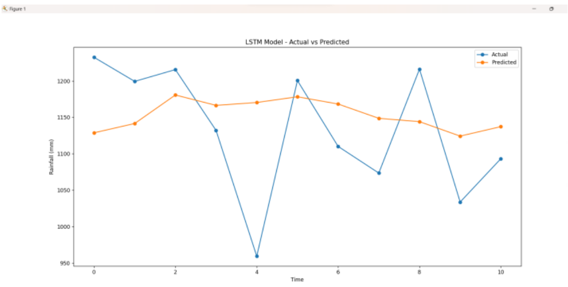
The input features, such as historical rainfall data, were carefully processed and fed into the LSTM network. The model underwent an iterative training process, adjusting its internal parameters to minimize the difference between its predictions and actual rainfall values.

Through this dynamic learning process, the LSTM model learned to discern complex patterns and temporal trends within the historical dataset, offering a powerful alternative approach to ensemble learning for rainfall prediction.

```
# Normalize the features using Min-Max scaling
scaler_X = MinMaxScaler()
X_scaled = scaler_X.fit_transform(X)

# Normalize the target variable using Min-Max scaling
scaler_y = MinMaxScaler()
y_scaled = scaler_y.fit_transform(y.values.reshape(-1, 1))

# Create sequences for the LSTM model
sequence_length = 12
X_lstm, y_lstm = [], []
for i in range(len(X_scaled) - sequence_length):
    X_lstm.append(X_scaled[i:i + sequence_length, :])
    y_lstm.append(y_scaled[i + sequence_length])

X_lstm, y_lstm = np.array(X_lstm), np.array(y_lstm)

# Split the data into training and testing sets
X_train_lstm, X_test_lstm, y_train_lstm, y_test_lstm = train_test_split(X_lstm, y_lstm, test_size=0.2, random_state=42)

# Build the LSTM model
model = Sequential()
model.add(LSTM(50, input_shape=(X_train_lstm.shape[1], X_train_lstm.shape[2])))
model.add(Dense(1))
model.compile(optimizer='adam', loss='mean_squared_error')

# Train the model
model.fit(X_train_lstm, y_train_lstm, epochs=50, batch_size=32, verbose=1)
```

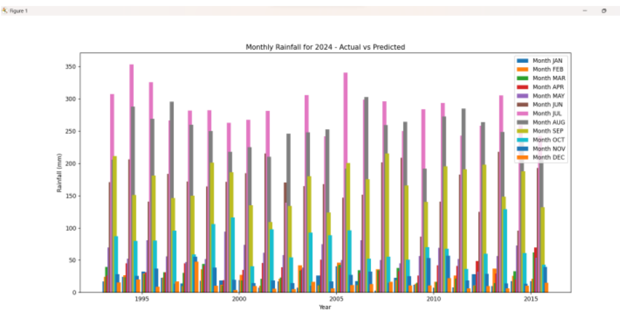
Result

LSTM Model prediction for 2024



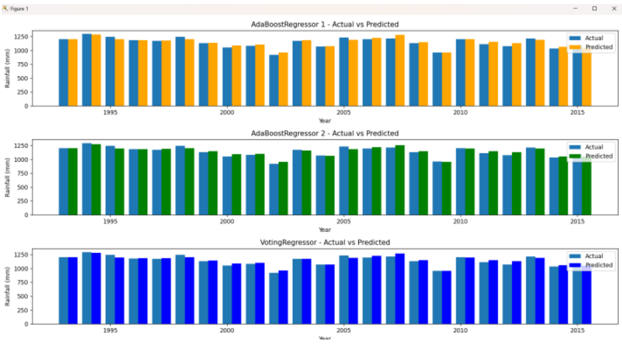

```
Epoch 49/50
3/3 [==============================] - 0s 3ms/step - loss: 0.0331
Epoch 50/50
3/3 [==============================] - 0s 3ms/step - loss: 0.0330
1/1 [==============================] - 0s 258ms/step

Evaluation Metrics for the Test Set:
Mean Absolute Error (MAE): 73.11217817826704
R-squared Score: -0.09769718619263768
```

Predicted rainfall for the year 2024 (in mm)



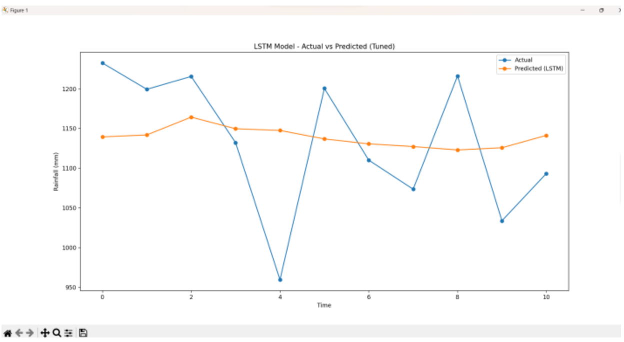
```
Evaluation Metrics for the Test Set:
Mean Absolute Error (MAE): 73.11217817826704
R-squared Score: -0.09769718619263768
Predicted Rainfall for Each Month i 2024:
JAN: 1195.7955320938102 mm
FEB: 1267.9562498047446 mm
MAR: 1209.3457784151306 mm
APR: 1181.9433951160456 mm
OCT: 920.8 mm
NOV: 1174.5 mm
DEC: 1071.3 mm

Predicted Seasonal Rainfall in 2024:
Jan-Feb: 1114.3040517210832 mm
Mar-May: 1196.938741977791 mm
Jun-Sep: 1055.1729417668446 mm
Oct-Dec: 1053.367205961077 mm
```

Compared model



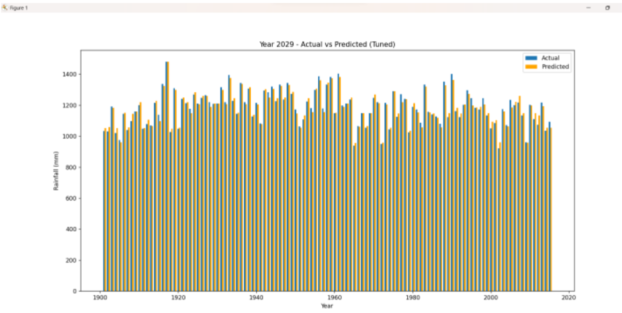

```
Evaluation Metrics for VotingRegressor:
Mean Absolute Error (MAE): 24.32941816971877
R-squared Score: 0.8915421858370806
```

LSTM Model prediction for 2029



```
Evaluation Metrics for the LSTM Model:
Mean Absolute Error (MAE): 70.81560280539772
R-squared Score: 0.014210523890344606
```

Predicted rainfall for the year 2024 (in mm)




```
Best Hyperparameters: {'learning_rate': 0.2, 'loss': 'square', 'n_estimators': 50}
Actual vs Predicted Rainfall for Each Month in 2029 (Tuned):
YEAR: Actual - 1201.9 mm, Predicted - 1203.2923076923075 mm
JAN: Actual - 1295.6 mm, Predicted - 1273.1307692307694 mm
FEB: Actual - 1243.6 mm, Predicted - 1197.1499999999999 mm
MAR: Actual - 1181.8 mm, Predicted - 1184.041935483871 mm
APR: Actual - 1171.4 mm, Predicted - 1188.6705882352942 mm
MAY: Actual - 1243.5 mm, Predicted - 1203.68 mm
JUN: Actual - 1132.0 mm, Predicted - 1147.0100000000004 mm
JUL: Actual - 1050.4 mm, Predicted - 1092.475 mm
AUG: Actual - 1083.3 mm, Predicted - 1101.925 mm
SEP: Actual - 920.8 mm, Predicted - 958.5333333333334 mm
OCT: Actual - 1174.5 mm, Predicted - 1158.474074074074 mm
NOV: Actual - 1071.3 mm, Predicted - 1061.7928571428572 mm
DEC: Actual - 1232.5 mm, Predicted - 1185.0249999999996 mm

Evaluation Metrics for the Entire Test Set (Tuned):
Mean Absolute Error (MAE): 25.293437593222926
R-squared Score: 0.893404314584308
```

Conclusion

In conclusion, this research endeavors to advance the precision of rainfall predictions in India by introducing a novel ensemble learning model that integrates AdaBoostRegressor, GradientBoostingRegressor, RandomForestRegressor, and VotingRegressor. Through an extensive evaluation process, we have demonstrated the efficacy of our model in forecasting rainfall for the next five years based on a robust 115-year dataset (1901-2015). Our findings reveal that the integration of diverse regression algorithms enhances predictive accuracy, showcasing the model's proficiency in capturing the complex and dynamic patterns of India's rainfall. The hyperparameter tuning and model evaluation processes have underscored the significance of these components in achieving optimal performance.

Furthermore, our research holds practical implications for diverse sectors, contributing valuable insights to the government of India. The ability to make accurate rainfall predictions is paramount for decision-making in critical areas such as Waste Water Treatment Plants and Hydroelectric energy reservoirs, both of which rely heavily on timely and precise precipitation information. By providing a more accurate and reliable forecasting tool, our ensemble learning model equips decision-makers with the information needed to enhance operational efficiency and resource management.

As with any scientific endeavor, our study is not without limitations. The success of our model hinges on the quality and representativeness of the historical data, and uncertainties in climate patterns may introduce challenges. Additionally, future work could explore further refinements in model architecture, incorporation of additional relevant features, and adaptation to changing climatic conditions.

In essence, this research serves as a stepping stone in the pursuit of improved rainfall prediction methodologies, contributing not only to the scientific understanding of ensemble learning applications but also to the practical realms of environmental and resource management. As we move forward, the insights gained from this study pave the way for continued advancements in weather forecasting, aligning with the global imperative of harnessing technology to address climate-related challenges.

References

[1] Deeksha Tripathi, e-Prime - Advances in Electrical Engineering, Electronics and Energy, Volume 6, December 2023, 100296.

[2] Mumtaz ali, Journal of Hydrology, Volume 584, May 2020, 124647.

[3] Sam Cramer, Expert Systems with Applications,
Volume 85, 1 November 2017, Pages 169-181.

[4] Boris Thies, Remote Sensing of Environment.Volume 141, 5 February 2014, Pages 129-143.

[5] Wei-Chiang Hong, Applied Mathematics and Computation,Volume 200, Issue 1, 15 June 2008, Pages 41-57.

[6] S.S. Badhiye, P.N. Chatur, B.V. Wakode, Temperature and humidity data analysis for future value prediction using clustering technique: an approach, Int. J. Emerg. Technol. Adv. Eng., 2 (1) (2012), pp. 88-91.

[7] D.K. Pabreja, Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst, Int. J. Comput. Sci. Inf. Technol. (IJCSIT), 3 (1) (2012), pp. 2996-2999.

[8] S. Kannan, S. Ghosh, Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output, Stoch. Environ. Res. Risk Assess., 25 (4) (2011), pp. 457-474.

[9] M.J. Page, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ, 372 (2021), p. n71.

[10] Kitchenham B. and Charters S. (2007) “Guidelines for performing systematic literature reviews in software engineering,” 2.

[11] S. Myint, Linear regressions of predicting rainfall over Kalay region, Int. J. Trend Sci. Res. Dev. (IJTSRD), 3 (5) (2019), pp. 2456-6470.

[12] J. Sulaiman, S.H. Wahab, Heavy rainfall forecasting model using artificial neural network for flood prone area, K.J. Kim, H. Kim, N. Baek (Eds.), IT Convergence and Security 2017, Springer Singapore, Singapore (2018), pp. 68-76.

[13] B.T. Pham, et al., A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area, Geotech. Geol. Eng., 34 (6) (2016), pp. 1807-1824.

[14] S. Cramer, et al., An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives, Expert Syst. Appl., 85 (2017), pp. 169-181.

[15] M. Kim, et al., Evaluation of the k-nearest neighbor method for forecasting the influent characteristics of wastewater treatment plant, Front. Environ. Sci. Eng., 10 (2) (2016), pp. 299-310.

[16] S. Zainudin, D.S. Jasim, A. Abu Bakar, Comparative analysis of data mining techniques for malaysian rainfall prediction, Int. J. Adv. Sci. Eng. Inf. Technol., 6 (6) (2016), pp. 1148-1153.

[17] N.S. Sani, et al., Ensemble learning for rainfall prediction, Int. J. Adv. Comput. Sci. Appl. (IJACSA), 11 (11) (2020), pp. 153-162.

[18] Y. Ren, L. Zhang, P.N. Suganthan, Ensemble classification and regression-recent developments, applications and future directions [review article], IEEE Comput. Intell. Mag., 11 (1) (2016), pp. 41-53.

[19] Z.H. Zhou, Ensemble Methods: Foundations and Algorithms, CRC Press (2012).

[20] P.S. Dutta, H. Tahbilder, Prediction of rainfall using data mining technique over Assam, Indian J. Comput. Sci. Eng. (IJCSE), 5 (2) (2014), pp. 85-90.

[21] S.Y. Ji, et al., Designing a rule-based hourly rainfall prediction model, Proceedings of the IEEE 13th International Conference on Information Reuse & Integration (IRI) (2012), pp. 303-308.

[22] N. Khandelwal, R. Davey, Climatic assessment of Rajasthan's region for drought with concern of data mining techniques, Concern, 2 (5) (2012), pp. 1695-1697.

[23] E. Kaya, B. Barutçu, Ş.S. Menteş, A method based on the Van der Hoven spectrum for performance evaluation in prediction of wind speed, Turk. J. Earth Sci., 22 (4) (2013), pp. 681-689.

[24] S. Shanmuganathan, P. Sallis, Data mining methods to generate severe wind gust models, Atmosphere (Basel), 5 (1) (2014), pp. 60-80.

[25] Amarakoon, ‘Effectiveness of using data mining for predicting climate change in Sri Lanka’ (2010), pp. 2321–8169.

[26] S. Navaz, H. Khan, S.M. Ghosh, A survey on ensemble computing method for rainfall prediction in different regions of Chhattisgarh, Int. J. Sci. Res. (IJSR), 6 (6) (2017), pp. 19-25.

[27] Petre, E.G., (2009). ‘A decision tree for weather prediction. Universitatea Petrol-Gaze din Ploiesti’, 61(1), pp.77–82.

[28] Kannan, M., Prabhakaran, S. and Ramachandran, P., (2010). ‘Rainfall forecasting using data mining technique’.

[29] P.J. Sallis, W. Claster, S. Hernández, A machine-learning algorithm for wind gust prediction, Comput. Geosci., 37 (9) (2011), pp. 1337-1344.