AUTOMATED WEATHER FORECASTING POWERED BY MACHINE LEARNING

Pradeepa S

Department of Artificial Intelligence

Kongu Engineering College

Erode, India

pradeepas.22aim@kongu.edu

Roshini G

Department of Artificial Intelligence

Kongu Engineering College

Erode, India

roshinig.22aim@kongu.edu

Yoganath M

Department of Artificial Intelligence

Kongu Engineering College

Erode, India

yoganathm.22aim@kongu.edu

***Abstract*--This project focuses on the development of a sophisticated weather forecasting system utilizing advanced machine learning methodologies. The system employs time series models, including Long Short-Term Memory (LSTM) networks and the Prophet model, for precise temperature forecasting. Additionally, it integrates regression algorithms such as Random Forest, Decision Tree, AdaBoost, and XGBoost to enhance temperature prediction accuracy. The dataset, "Seattle Weather.csv," containing 1,462 entries with variables like wind speed, temperature (min/max), and precipitation, is analyzed. For weather classification, the system utilizes algorithms like Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and XGBoost. To optimize model performance, techniques such as oversampling and temporal resampling to monthly averages are applied. The system's performance is rigorously evaluated using accuracy metrics and predictive capabilities.**

**Keywords:** Time series forecasting, LSTM networks, prophet model, machine learning, regression model, classification model, and data resampling.

# **I.INTRODUCTIO****N**

Weather forecasting is critical to many industries, including disaster relief and agriculture. In order to transform weather forecasting, this initiative integrates cutting-edge machine learning approaches with conventional meteorological methodologies. Our method attempts to build a strong forecasting framework by combining time series analysis with regression and classification models. Using the "Seattle Weather Dataset," which includes critical characteristics such as temperature and precipitation, our method classifies weather patterns using classification models and uses regression models to estimate temperatures with accuracy. For precise future temperature projections, we also use time series models like LSTM and Prophet models.

To enhance model accuracy, we apply oversampling techniques for both regression and classification tasks. This approach addresses data imbalance, ensuring that underrepresented classes are better reflected in the model and improving overall performance. Additionally, we utilize resampling methods, such as averaging monthly data and optimizing training time on historical data to further refine prediction accuracy and efficiency.

Our primary goal is to elevate the efficiency and precision of weather forecasting through data-driven technology. By combining machine learning techniques, oversampling for classification and regression, and resampling for time series analysis, we aim to enhance weather intelligence. Our efforts focus on improving accuracy, predictive power, and model performance in weather forecasting.

# **II.LITERATURE SURVEY**

Weather forecasting has been greatly enhanced by recent developments in machine learning (ML), increasing its accuracy and dependability for a range of uses. Farman et al. (2024) investigated how to combine machine learning (ML) with pre-existing forecasting models to enhance precipitation forecasts in Pakistan. Using the K-Nearest Neighbor (KNN) algorithm, they were able to obtain 84% accuracy. This paper demonstrates how the analysis of complex data linkages by AI-based models can outperform classical methods. Calvo-Olivera et al. (2024) employed a Random Forest (RF) classifier to evaluate the uncertainty of weather forecasts. Their findings demonstrated that RF was a highly successful method for predicting precipitation based on strong performance measures. Their work emphasizes how crucial sophisticated machine learning models are to industries like water management and agriculture. In short-term weather forecasting, Kaur and Kumar (2024) achieved an amazing accuracy of 97.60% using Random Forest, demonstrating the model's effectiveness in collecting and predicting dynamic weather variables. AbdulRahim et al. (2022) assessed a number of machine learning classifiers for weather prediction and discovered a decision tree method that was exceptionally successful, achieving an optimal accuracy of 100%. This finding emphasizes the significance of choosing an appropriate algorithm based on a specific dataset. When Apaydin et al. (2022) compared several machine learning algorithms for temperature forecasting, XGBoost and LightGBM stood up as the best options due to their excellent R2 scores and ability to handle complex meteorological data with resilience. Gradient boosted regression was used by Singh et al. (2021) to forecast wind power and obtained an R2 value of 0.9651, indicating the model's efficacy in incorporating forecasts of renewable energy sources into smart grid settings. Lastly, Liyew and Melese (2021) employed ML models like XGBoost and environmental features to anticipate daily precipitation. Their method was shown to be highly accurate when compared to previous approaches, and they proposed that adding sensor data could increase the prediction accuracy even further. When combined, these studies show how machine learning has revolutionized weather forecasting, demonstrating how it can increase prediction accuracy and efficiently address challenging weather issues.

# **III.DATA DESCRIPTION**

The primary objective of this study is to evaluate the efficacy of predictive models for regional weather conditions. This analysis employs a Kaggle-sourced dataset comprising 1,461 records, each featuring six key meteorological attributes essential for accurate weather forecasting. The dataset facilitates a rigorous assessment of model performance across various forecasting algorithms (TABLE 1).

The dataset includes the following attributes:

**Date**: Represents the date of the recorded weather data.

**Precipitation**: Quantifies the amount of precipitation observed.

**Temp\_max**: Indicates the maximum temperature recorded.

**Temp\_min**: Represents the minimum temperature recorded.

**Wind**: Provides information on the recorded wind speed.

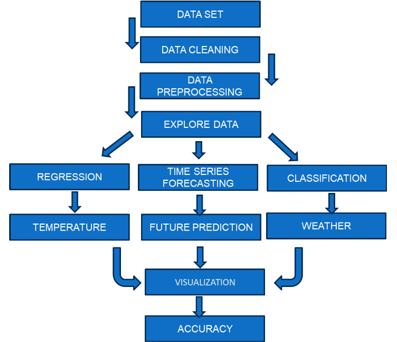
**Weather**: Represents a categorical variable describing

the overall weather conditions.

|  |  |  |  |
| --- | --- | --- | --- |
| S NO | COLUMN | NON-NULL COUNT | D\_TYPE |
| 1 | Date | 1461 non-null | object |
| 2 | Precipitation | 1461 non-null | float64 |
| 3 | Temp\_max | 1461 non-null | float64 |
| 4 | Temp\_min | 1461 non-null | float64 |
| 5 | Wind | 1461 non-null | float64 |
| 6 | Weather | 1461 non-null | object |

**TABLE 1: Data Distribution**

# **IV.MODEL ARCHITECTURE**



**IV.PROPOSED SYSTEM**

The planning process is divided using 3 types of models to view behavior and reality: **A.** Distribution models, **B.** Regression models, **C.** Time series forecasting models. We use various algorithms to accurately measure weather on every project.

***A.* PREPROCESSING**

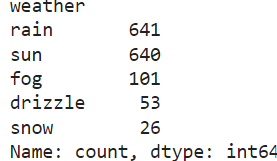
Feature selection plays a pivotal role in the development of effective machine learning models by identifying and utilizing the most relevant features from a dataset. This process reduces computational complexity by eliminating extraneous and inconsistent data. In our approach, we emphasize the importance of transforming date-related information into more meaningful features.

Specifically, the date attribute is decomposed into its constituent parts—week, month, and year—which are then converted into binary values using the Label Encoder. This encoding method ensures that the temporal aspects of the data are represented in a format that is more interpretable and useful for machine learning algorithms, leading to a more streamlined and efficient data processing workflow. By focusing on the most informative features, our approach enhances the overall performance and interpretability of the resulting models (TABLE 2).

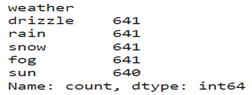
|  |  |  |  |
| --- | --- | --- | --- |
| S\_NO | COLUMN | NON-NULL COUNT | D\_TYPE |
| 1 | Day-of-week | 1461 non-null | Int64 |
| 2 | Month | 1461 non-null | Int64 |
| 3 | Year | 1461 non-null | Int64 |

**TABLE 2: Data Distribution**

In our dataset the class values are imbalanced way



To address class imbalance, oversampling was employed to ensure equal representation across the rain, sun, fog, and snow categories. By artificially increasing the minority class samples, this technique mitigates bias and enhances the robustness and accuracy of the model, providing a more balanced distribution of class values during training.



Preprocessing in time series forecasting involves converting date columns to datetime format and setting them as index values. Period-based data aggregation, such as calculating monthly averages, supports trend analysis and improves model training.

**B. I. CLASSIFICATION MODELS**

In the weather forecast classification study, the input features include precipitation, minimum temperature, maximum temperature, wind speed, day of the week, month, and year. The following models are utilized:

**1.Decision Tree:** The Decision Tree algorithm is utilized for its ability to effectively partition the feature space and capture non-linear relationships within the climate data. It is adept at providing insights into feature interactions and decision paths. However, to avoid overfitting especially in complex datasets careful hyperparameter tuning and pruning techniques are required to achieve reliable performance.

**2.Random Forest:** The Random Forest algorithm, an ensemble method, is used to improve forecast accuracy and model robustness. By combining multiple decision trees, each trained on different subsets of the data, Random enhances the model's ability to generalize to unseen data. This approach is particularly effective for climate data, as it captures complex patterns and interactions among features like precipitation, temperature, wind speed, and temporal elements.

3.**XGBoost:** XGBoost (Extreme Gradient Boosting) is employed to enhance forecasting accuracy by capturing complex patterns in climate data. This gradient boosting framework excels at modeling intricate feature interactions and requires careful hyperparameter tuning—such as adjusting the learning rate, number of boosting rounds, and tree depth—to optimize performance. It also demands significant computational resources and efficient hardware for handling large datasets and complex model configurations.

4.**KNN:** K-Nearest Neighbors (KNN) is used for its simplicity and effectiveness in detecting clusters within weather data. It classifies data based on proximity to nearest neighbors, making it suitable for identifying patterns in the feature space. However, KNN is sensitive to the choice of distance metrics and can be affected by noisy or irrelevant features. Proper feature selection and preprocessing are crucial to enhance its performance and accuracy.

These models were chosen for their ability to maintain relationships, minimize overfitting, and accurately describe climates based on given characteristics.

**B. II. EVALUATION METRICS**

**1. Accuracy:** Accuracy measures the proportion of correctly classified instances among the total number of instances. It is defined as:

This metric indicates the overall correctness of the model's predictions.

**2. Precision**: Precision evaluates the proportion of true positive predictions out of all positive predictions made by the model. It is calculated as:

High precision signifies that the model's positive predictions are mostly correct.

**3. Recall:** Recall measures the proportion of actual positive instances that are correctly identified by the model. The formula in

High recall indicates that the model captures most of the positive instances.

**4.** **F1 Score:** The F1 Score combines precision and recall into a single metric, providing a balance between the two. It is calculated using:

This metric is useful when both precision and recall are important.

**5.** **ROC AUC Score:** The ROC AUC Score is derived from the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR). A higher AUC score indicates better model performance in class discrimination. The formula for the ROC AUC Score itself is:

ROC AUC Score = Area Under the ROC Curve

True Positive Rate (TPR), also known as Recall, is given by:

False Positive Rate (FPR) is given by:

6. **Log Loss (Logarithmic Loss):** Log Loss evaluates the performance of a classification model by penalizing incorrect predictions based on their probabilities. It is given by:

where yi is the actual label and pi is the predicted probability. Lower log loss indicates better model accuracy.

7. **Confusion Matrix:** The confusion matrix provides a detailed view of the model’s performance by showing the number of true positive, true negative, false positive, and false negative predictions:

This matrix helps in understanding where the model is making errors.

**8**. **Classification Report:** The classification report presents precision, recall, F1 score, and support for each class, offering a detailed evaluation of the model’s performance across different classes. These metrics provide a comprehensive view of the model’s accuracy and effectiveness, highlighting strengths and identifying areas for improvement.

**B. III. EXPERIMENTAL REUSLT**

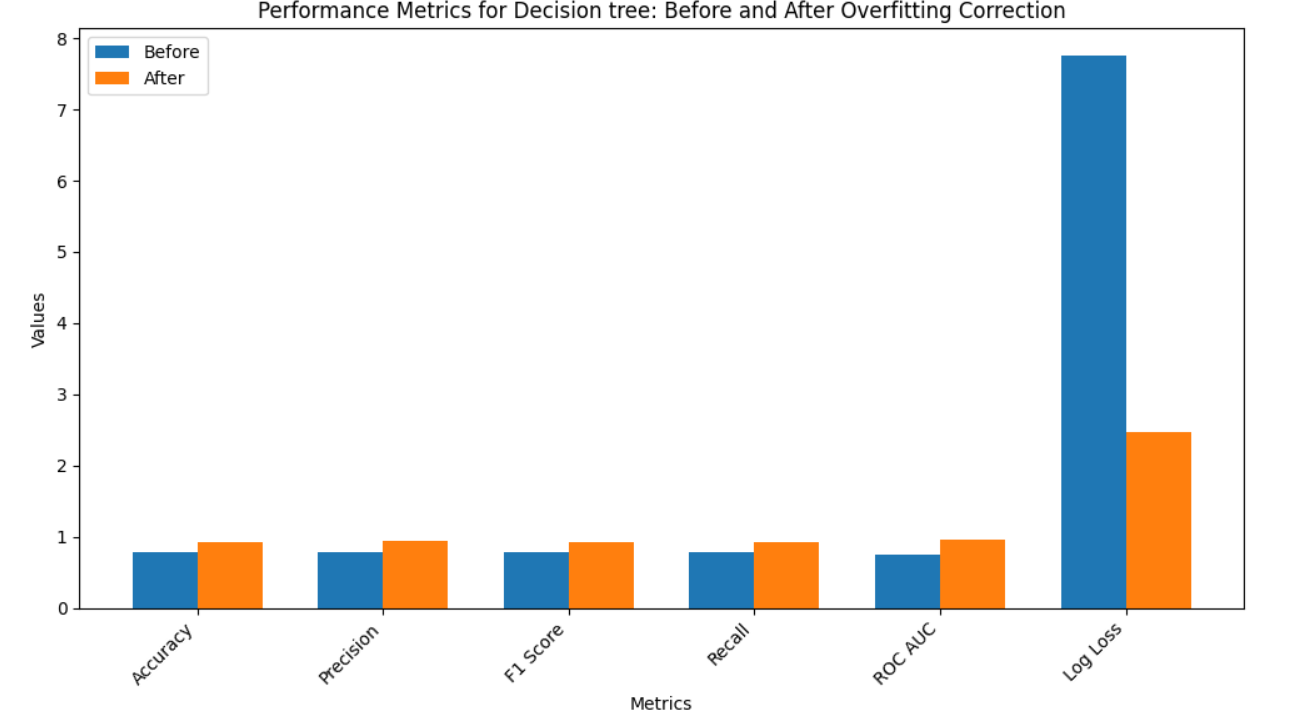
**Before:** Represents the algorithm's performance metrics obtained before implementing any overfitting mitigation techniques, reflecting its initial behaviour on the dataset.

**After:** Refers to the performance metrics of the algorithm following the application of methods to reduce overfitting, demonstrating its improved ability to generalize to new data

**DECISION TREE**

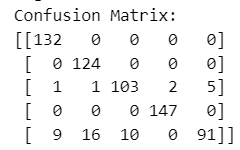
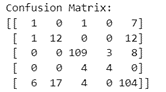
|  |  |  |
| --- | --- | --- |
| METRICS | BEFORE | AFTER |
| ACCURACY | 0.78498 | 0.93135 |
| PRECISION | 0.78801 | 0.93333 |
| F1 SCORE | 0.78498 | 0.93135 |
| RECALL | 0.7861 | 0.92835 |
| ROC AUC | 0.74760 | 0.95566 |
| LOG LOSS | 7.75000 | 2.47413 |

The above table shows significant performance gains after overfitting correction: accuracy increased from 0.78498 to 0.93135, precision from 0.78801 to 0.93333, and Log Loss decreased from 7.75000 to 2.47413, indicating improved model effectiveness and stability.

****

CONFUSION MATRIX

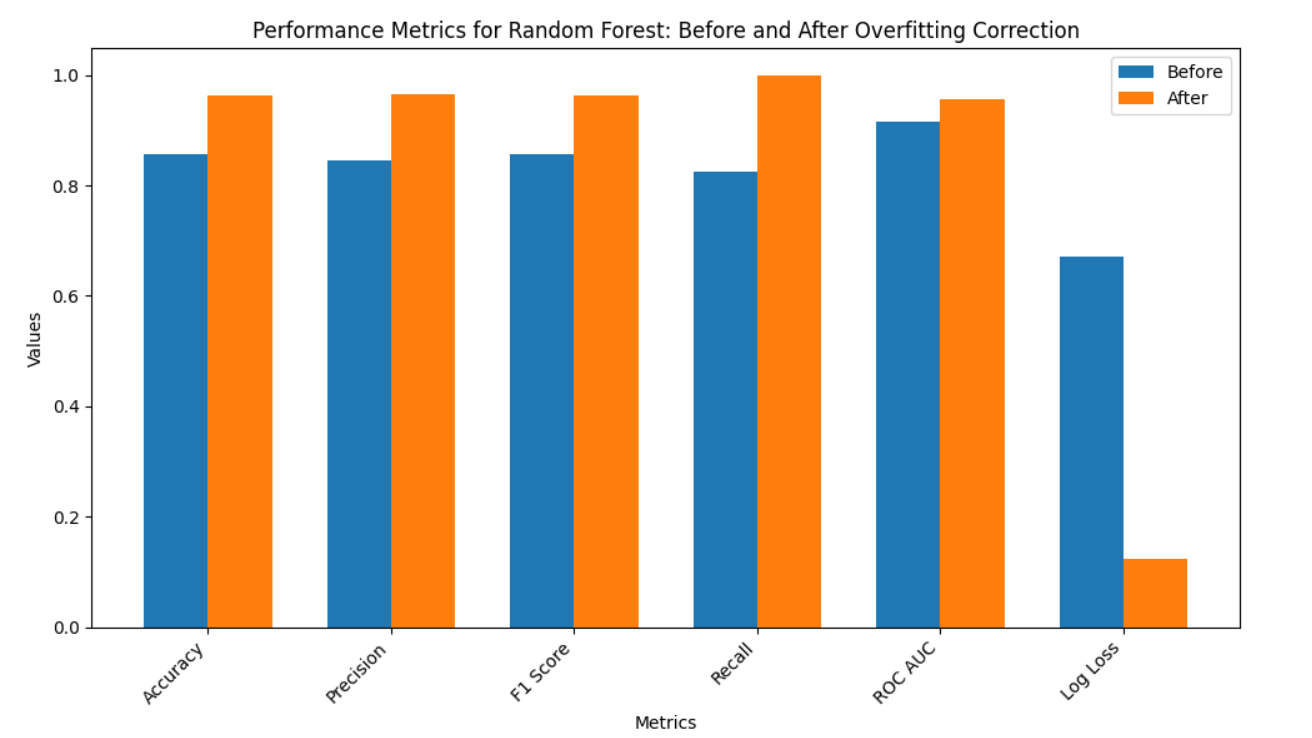
BEFORE AFTER



**RANDOM FOREST**

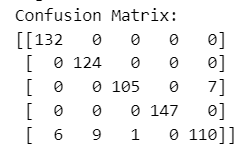
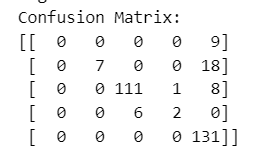
|  |  |  |
| --- | --- | --- |
| METRICS | BEFORE | AFTER |
| ACCURACY | 0.85665 | 0.96411 |
| PRECISION | 0.84491 | 0.96547 |
| F1 SCORE | 0.85665 | 0.96362 |
| RECALL | 0.82530 | 0.99856 |
| ROC AUC | 0.91633 | 0.12368 |
| LOG LOSS | 0.67141 | 0.12368 |

The above table shows Random Forest performance improvements after overfitting correction: accuracy increased from 0.85665 to 0.96411, precision from 0.84491 to 0.96547, and recall from .0.82530 to 0.99856. Log Loss decreased from 0.67141 to 0.12368, indicating better model effectiveness.



CONFUSION MATRIX

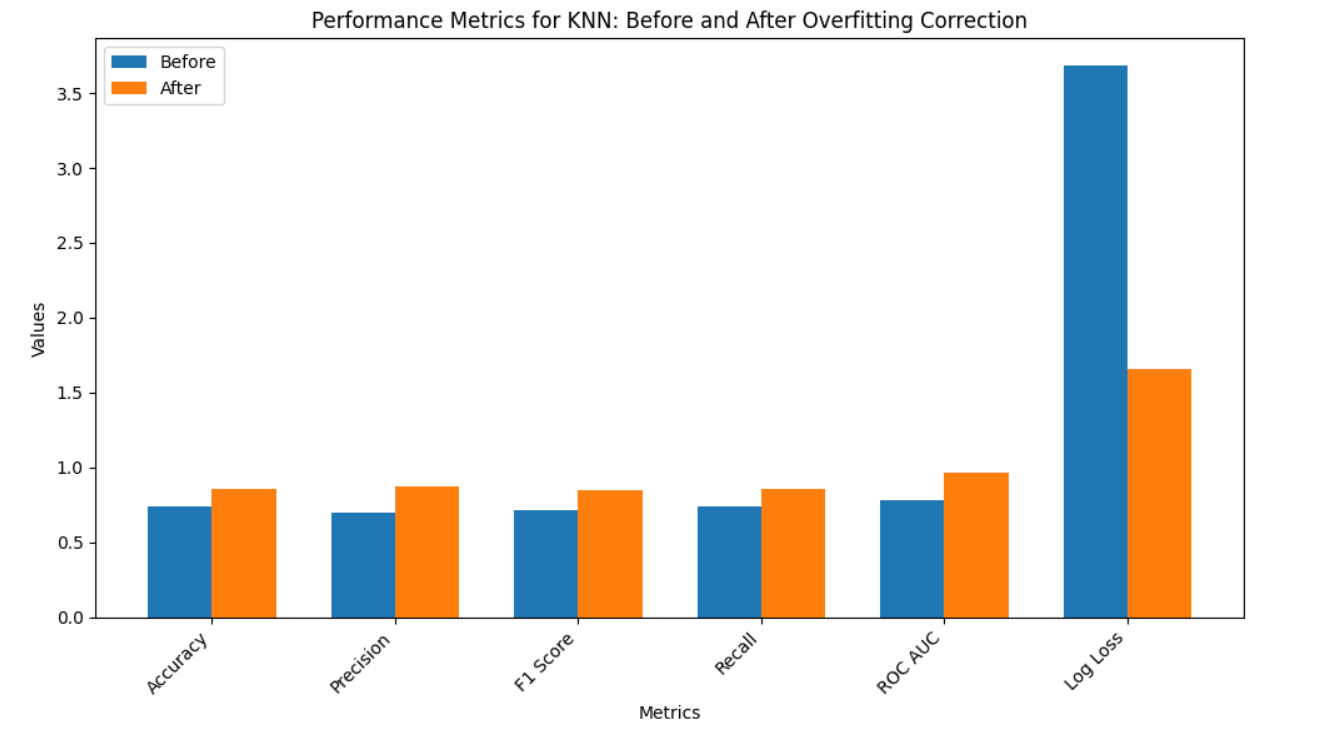
BEFORE AFTER



**KNN**

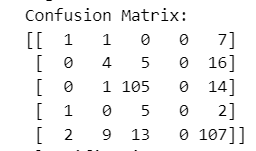
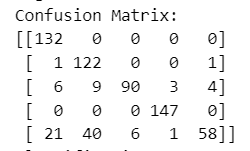
|  |  |  |
| --- | --- | --- |
| METRICS | BEFORE | AFTER |
| ACCURACY | 0.74061 | 0.85647 |
| PRECISION | 0.69406 | 0.87593 |
| F1 SCORE | 0.71400 | 0.84428 |
| RECALL | 0.74061 | 0.85647 |
| ROC AUC | 0.78497 | 0.96291 |
| LOG LOSS | 3.68163 | 1.65591 |

The above table shows significant performance gains after overfitting correction: accuracy increased from 0.74061 to 0.85647, Precision from 0.69406 to 0.87593, F1 Score from 0.71400 to 0.84428, and ROC AUC from 0.78497 to 0.96291. Log Loss decreased from 3.68163 to 1.65591, indicating better probability estimation.



CONFUSION MATRIX

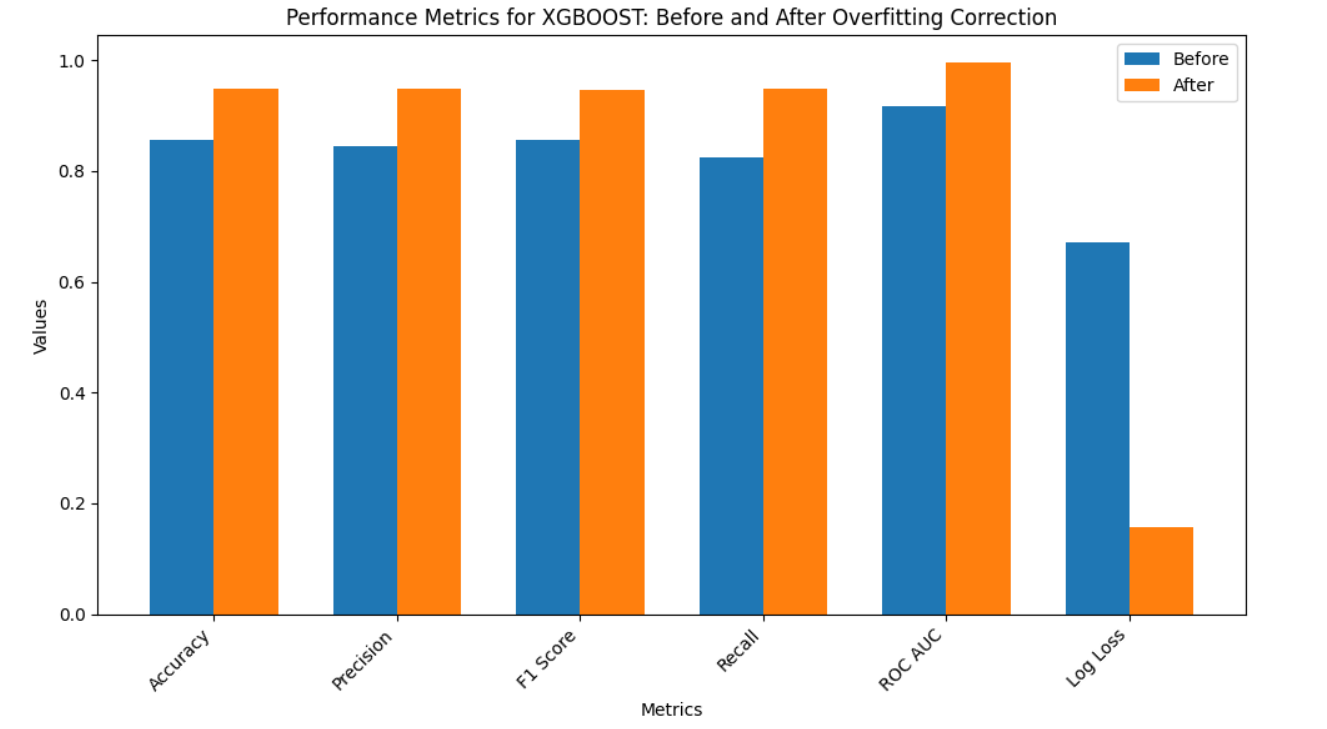
BEFORE AFTER

**XGBOOST**

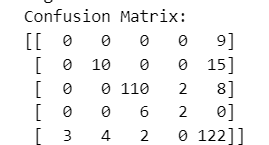
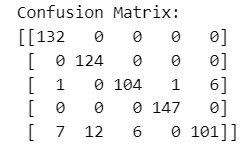
|  |  |  |
| --- | --- | --- |
| METRICS | BEFORE | AFTER |
| ACCURACY | 0.85665 | 0.94851 |
| PRECISION | 0.84491 | 0.94906 |
| F1 SCORE | 0.85665 | 0.94706 |
| RECALL | 0.82530 | 0.94851 |
| ROC AUC | 0.91633 | 0.99488 |
| LOG LOSS | 0.67141 | 0.15594 |

The above table shows the model's performance enhancements after overfitting correction: accuracy improved from 0.85665 to 0.94851, precision from 0.84491 to 0.94906, and recall from 0.82530 to 0.94851. Log Loss decreased significantly from 0.67141 to 0.15594, indicating a more effective model**.**

****

CONFUSION MATRIX

BEFORE AFTER

Among the evaluated models, Random Forest Classification demonstrated superior performance across multiple evaluation metrics. It achieved the highest accuracy, precision, and recall values, coupled with a substantial reduction in Log Loss, underscoring its effectiveness and reliability. These results indicate that Random Forest Classification is the most suitable model for our use case. Consequently, it was selected as the optimal model and has been preserved for ongoing and future applications.

**C. I. REGRESSION MODELS**

For our temperature function, we estimate the maximum temperature (Temp\_max) using parameters such as precipitation, temperature, wind speed, weather, week, month and year. Each algorithm has unique advantages:

**1.Random Forest Regression**: Utilizes an ensemble of decision trees to capture complex relationships in temperature data. It excels in handling high-dimensional data and mitigates overfitting through averaging. Its versatility allows it to address both regression and classification tasks effectively.

**2. Linear Regression:** A fundamental statistical model that estimates the relationship between the dependent variable and one or more independent variables using a linear equation. Linear regression is efficient for numerical data and requires minimal preprocessing, making it straightforward for understanding the impact of each feature on the target variable.

**3**.**AdaBoost Regression:** Enhances model accuracy by iteratively adjusting the weights of samples based on prediction errors. This method combines multiple weak learners to form a strong learner, improving robustness and generalization through sequential corrections and weighted sample emphasis.

**4.XGBoost Regression**: Advanced gradient boosting algorithm optimizes bias and variance to achieve precise temperature. It is a good calculation and suitable for big data with different characteristics. XGBoost regression is well received for its scalability, speed, and ability to handle missing elements.

**5.Decision Tree Regression:** A nonlinear model that provides interpretability by visualizing the decision paths impacting temperature predictions. It processes both numerical and categorical data, offering simplicity and the ability to reveal uncorrelated relationships without requiring extensive preprocessing.

**C. II. EVALUATION METRICS**

**1. Mean Absolute Error (MAE):** Measures the average absolute difference between predicted and actual values. It provides a straightforward indication of prediction accuracy, with lower values signifying better performance.

**2.Mean Squared Error (MSE**): Calculates the average of the squared differences between predicted and actual values, with lower values indicating better accuracy.

**3**. **Root Mean Squared Error (RMSE):** Measures the square root of the average of squared differences between predicted and actual values. It provides a measure of how spread out the residuals are, with lower values indicating better fit.

4. **R-Squared (R2):** Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, with higher values suggesting better model performance.

**C.III.EXPERIMENTAL RESULTS**

**Before:** Represents the algorithm's performance metrics obtained before implementing any overfitting mitigation techniques, reflecting its initial behaviour on the dataset.

**After:** Refers to the performance metrics of the algorithm following the application of methods to reduce overfitting, demonstrating its improved ability to generalize to new data

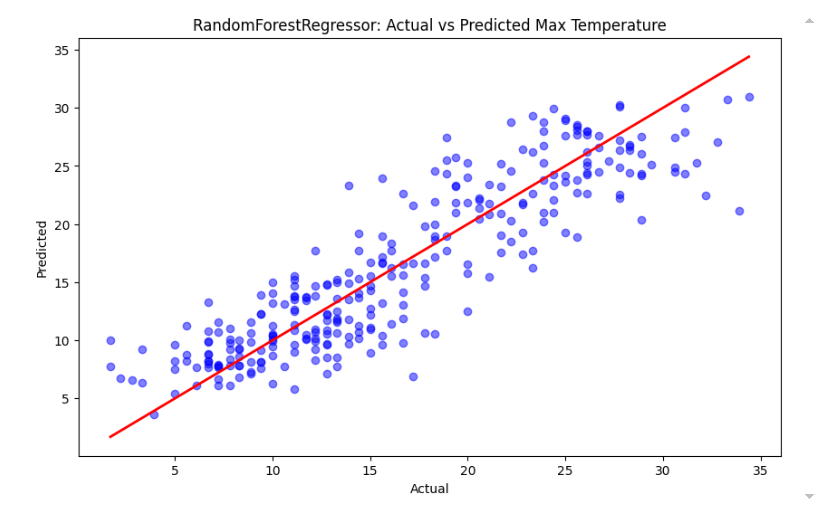
**RANDOM FOREST**

|  |  |  |
| --- | --- | --- |
| METRICS | BEFORE | AFTER |
| R2 SCORE | 0.7832 | 0.9294 |
| RMSE | 3.5127 | 2.1943 |
| MSE | 12.339 | 4.8150 |
| MAE | 2.7877 | 1.0648 |

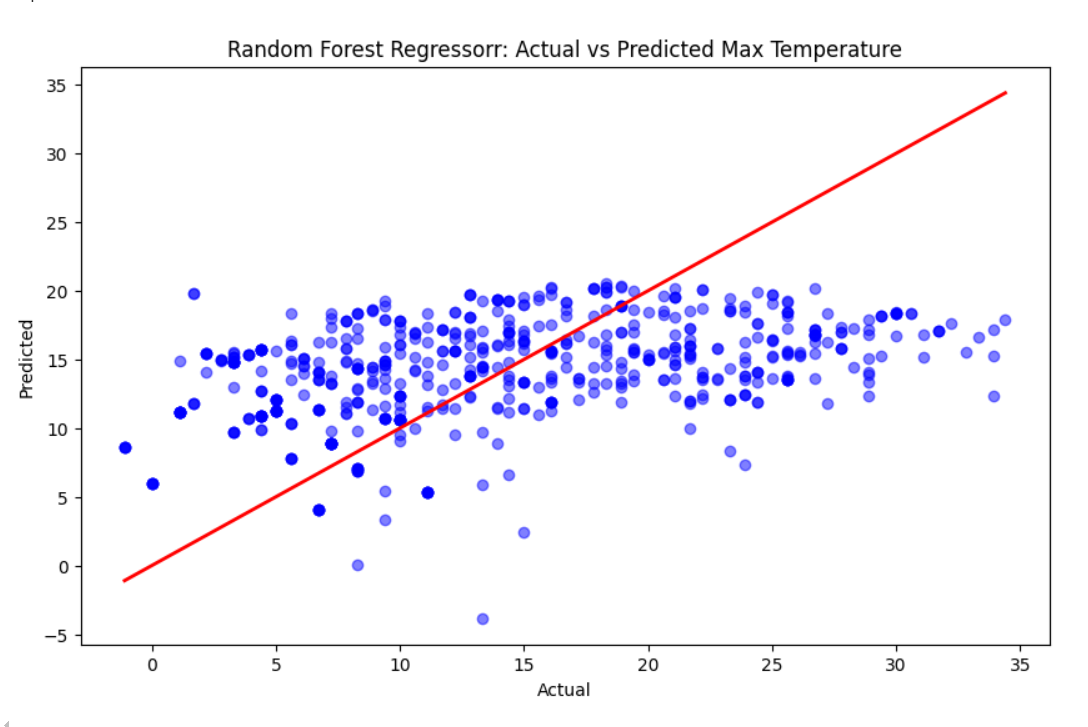
The above table shows that after overfitting correction, the R² score increased from 0.7832 to 0.9294, RMSE decreased from 3.5127 to 2.1943, MSE reduced from 12.339 to 4.8150, and MAE dropped from 2.7877 to 1.0648, indicating improved model accuracy and reduced error rates.

GRAPHICAL REPERSENTATION

BEFORE:



AFTER:



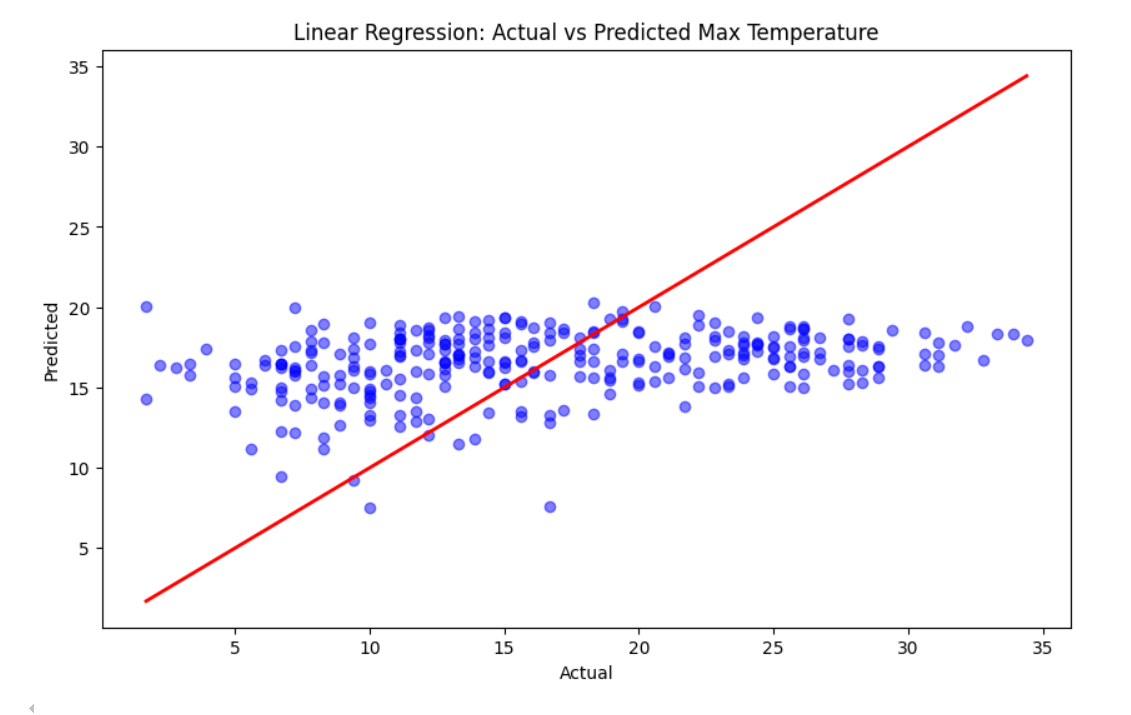
**LINEAR REGRESSION**

|  |  |  |
| --- | --- | --- |
| METRICS | BEFORE | AFTER |
| R2 SCORE | 0.0880 | 0.1741 |
| RMSE | 7.2040 | 7.5044 |
| MSE | 51.897 | 56.316 |
| MAE | 6.1172 | 6.2800 |

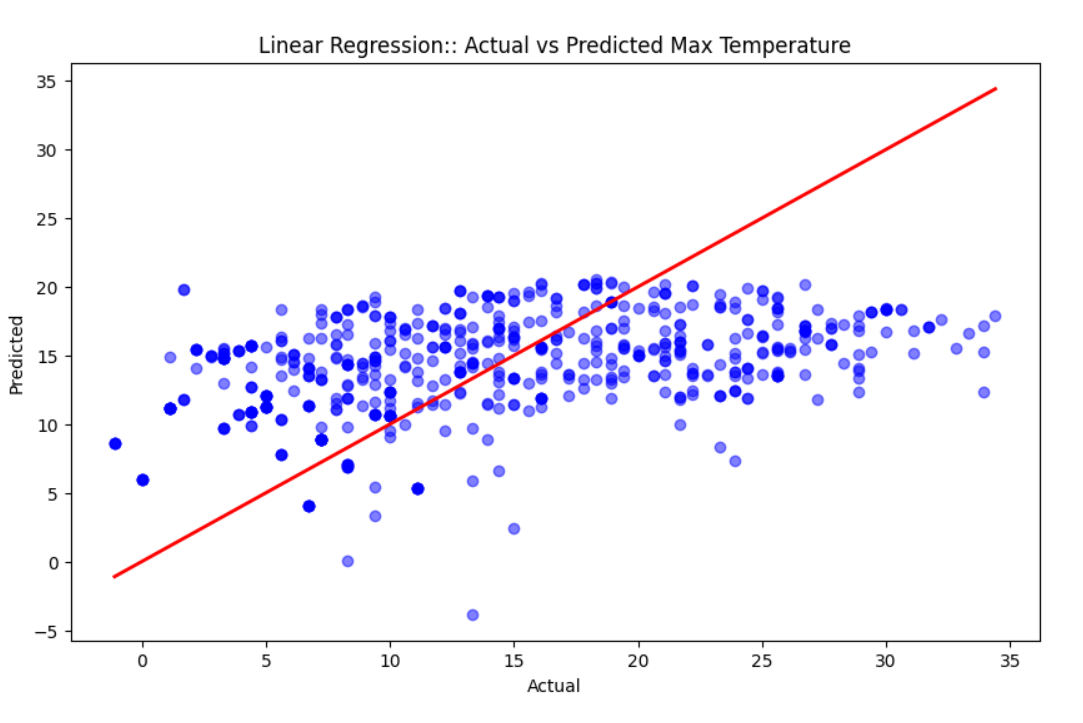
The above table highlights the regression metrics before and after overfitting correction. The R² score improved from 0.0880 to 0.1741, but the error metrics—RMSE, MSE, and MAE—increased slightly, indicating a trade-off between model fit and predictive accuracy.

GRAPHICAL REPERSENTATION

BEFORE:



AFTER:



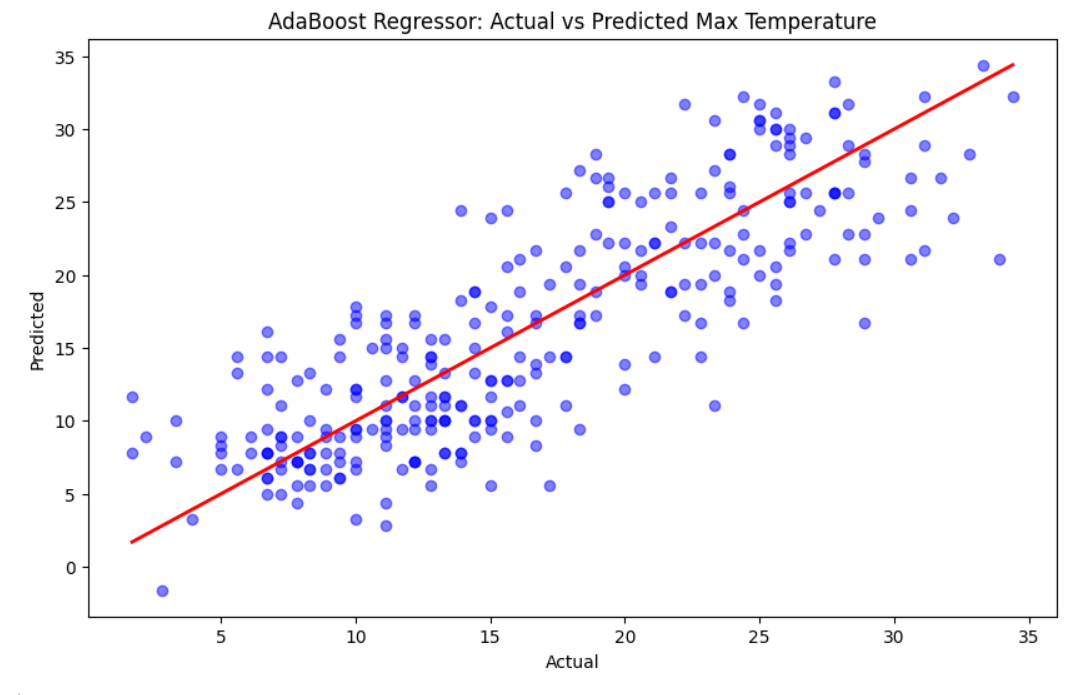
**ADABOOST REGRESSION**

|  |  |  |
| --- | --- | --- |
| METRICS | BEFORE | AFTER |
| R2 SCORE | 0.6716 | 0.7630 |
| RMSE | 4.3233 | 4.0201 |
| MSE | 18.6909 | 16.1610 |
| MAE | 3.3772 | 3.2895 |

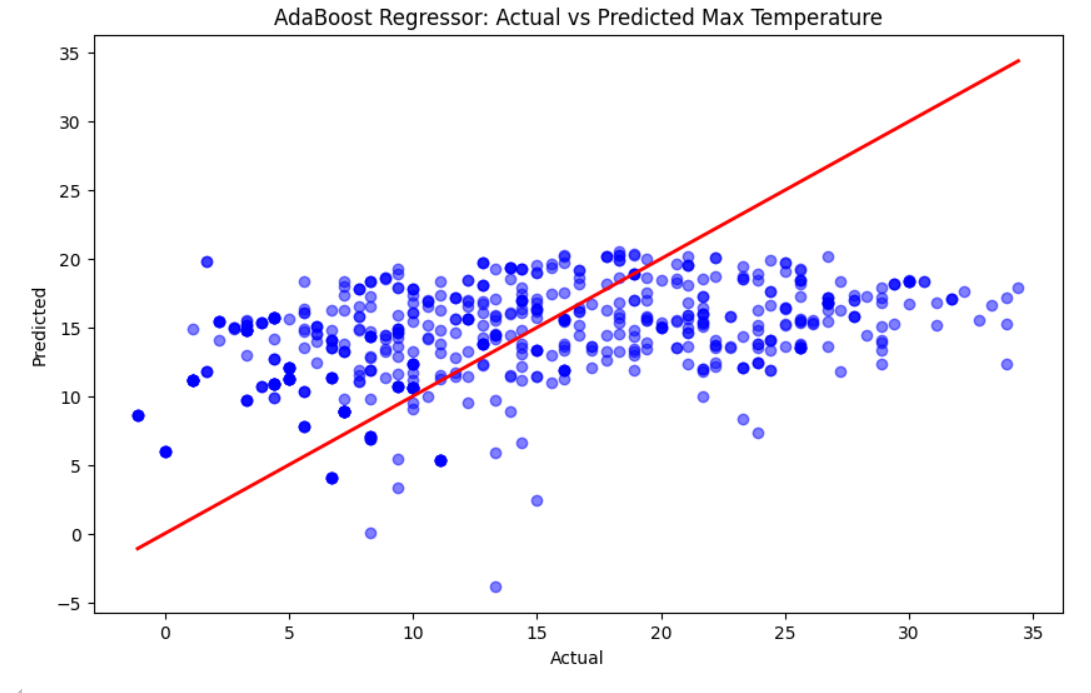
The above table demonstrates performance improvements after model tuning, with the R² score rising from 0.6716 to 0.7630. The reduction in RMSE from 4.3233 to 4.0201, MSE from 18.6909 to 16.1610, and MAE from 3.3772 to 3.2895 further indicates enhanced prediction accuracy and reduced errors in the model.

GRAPHICAL REPERSENTATION

BEFORE:



AFTER:



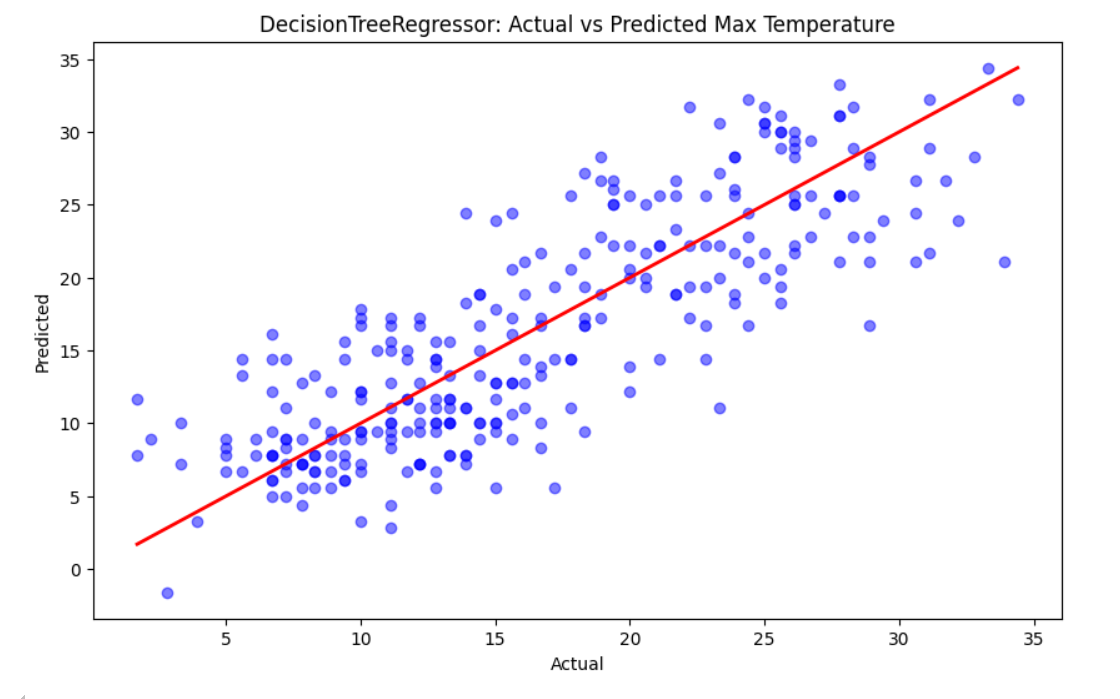
**DECISION TREE**

|  |  |  |
| --- | --- | --- |
| METRICS | BEFORE | AFTER |
| R2 SCORE | 0.6304 | 0.8869 |
| RMSE | 4.5862 | 2.777 |
| MSE | 21.0332 | 7.7119 |
| MAE | 3.7461 | 0.3317 |

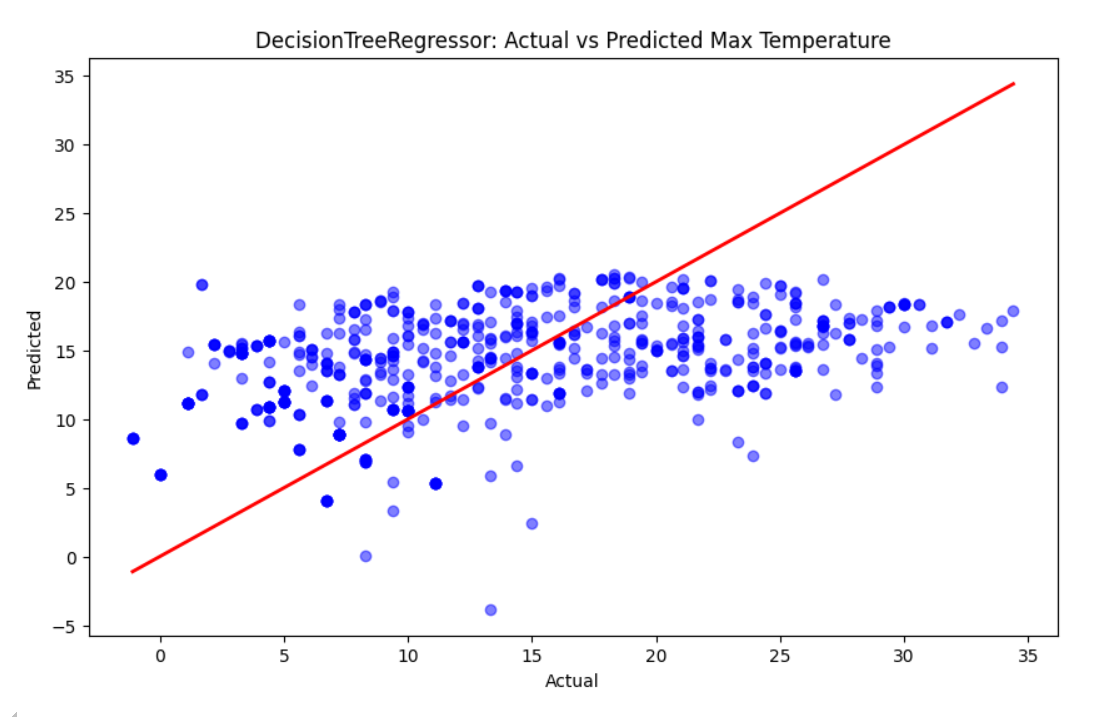
The table above shows significant improvements in model performance after tuning, with the R² score increasing from 0.6304 to 0.8869. Additionally, RMSE decreased from 4.5862 to 2.777, MSE from 21.0332 to 7.7119, and MAE from 3.7461 to 0.3317, reflecting better accuracy and lower prediction errors.

GRAPHICAL REPERSENTATION

BEFORE:



AFTER:



**D. I. TIMESERIES FORECASTING MODELS**

To forecast the maximum temperature (Temp\_max), we utilize features including precipitation, temperature, wind speed, weather conditions, week, month, and year. This approach allows us to predict temperature trends over various periods, leveraging historical data to enhance accuracy.

**LSTM:** Long Short-Term Memory (LSTM) networks, a specialized form of Recurrent Neural Networks (RNNs), are adept at time series forecasting due to their ability to capture and remember long-term dependencies. By leveraging their unique gating mechanisms, LSTMs maintain relevant information from historical data, allowing them to provide precise and reliable predictions for future outcomes in real-time applications.

**PROPHET:** The Prophet model, developed by Facebook, excels in forecasting time series data with strong seasonal patterns and multiple seasonal effects. It decomposes time series into components like trend, seasonality, and holidays, handling missing data and outliers effectively. By combining trend components with seasonality and holiday effects, Prophet delivers robust and interpretable forecasts for complex temporal patterns.

**D. II. EVALUATION METRICS**

**1. Mean Absolute Error (MAE):** Measures the average absolute difference between predicted and actual values. It provides a straightforward indication of prediction accuracy, with lower values signifying better performance.

**2.Mean Squared Error (MSE**): Calculates the average of the squared differences between predicted and actual values, with lower values indicating better accuracy.

**3**. **Root Mean Squared Error (RMSE):** Measures the square root of the average of squared differences between predicted and actual values. It provides a measure of how spread out the residuals are, with lower values indicating better fit.

4. **R-Squared (R2):** Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, with higher values suggesting better model performance.

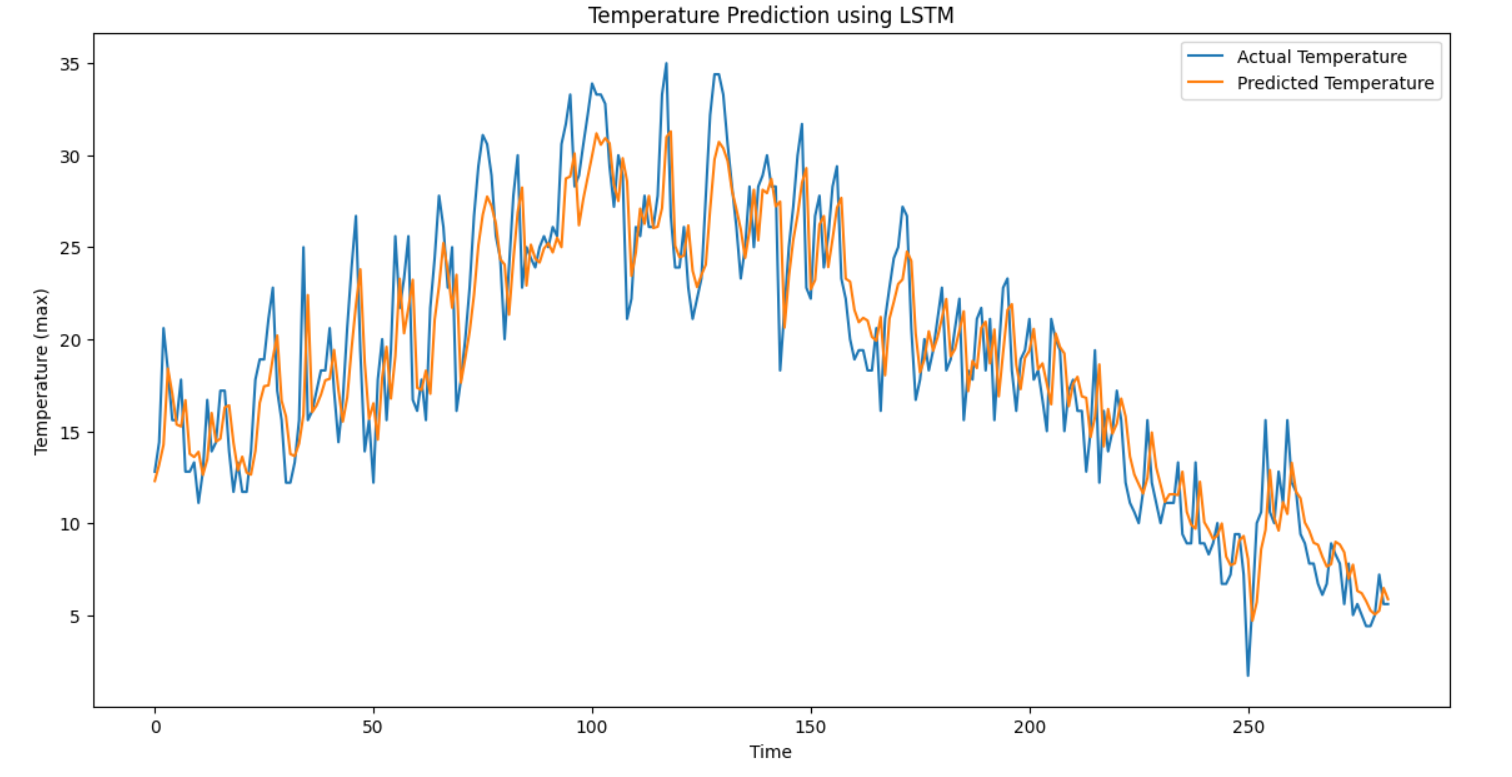
5. **Mean Absolute Percentage Error (MAPE)**: Measures the size of errors as a percentage of the actual values, providing a normalized measure of prediction accuracy.

**D.III.EXPERIMENTAL RESULTS**

**LSTM**

|  |  |
| --- | --- |
| METRICS | BEFORE |
| R2 SCORE | 0.6304 |
| RMSE | 4.5862 |
| MSE | 21.0332 |
| MAE | 3.7461 |

The table presents the model’s performance metrics prior to optimization. The R² score was 0.6304, indicating the proportion of variance explained by the model. The RMSE was 4.5862, representing the average deviation of the predictions from the actual values. The MSE was 21.0332, highlighting the squared errors average. The MAE was 3.7461, showing the average absolute error in predictions. These metrics reflect the model’s accuracy and error levels before any enhancements were applied.



**LSTM**

|  |  |
| --- | --- |
| METRICS | BEFORE |
| R2 SCORE | 0.6304 |
| RMSE | 4.5862 |
| MSE | 21.0332 |
| MAE | 3.7461 |

# **V.CONCLUSION**

The climate model developed for this study has proven to be highly effective in many applications. Random Forest is the best in classification results with 96% accuracy, proving its ability to handle complex data for cloud classification. Random forest regression performed well on temperature, with an impressive 94% accuracy rate for the regression function. Auto\_arima and SARIMA models have the same 92% accuracy over the forecast period, demonstrating their effectiveness in identifying seasonal patterns and trends. By combining different models, the accuracy of the model is further improved and the data representation is balanced across classes. Taken together, the overview shows how machine learning can transform climate change through accurate and reliable predictions.

# **VI.REFERENCES**

1. Chiranjeevi, B. S., et al. "Weather Prediction Analysis using Classifiers and Regressors in Machine Learning." 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2023.
2. Anusha, N., M. Sai Chaithanya, and Guru Jithendranath Reddy. "Weather prediction using multi linear regression algorithm." IOP Conference Series: Materials Science and Engineering. Vol. 590. No. 1. IOP Publishing, 2019.
3. Karna, Neha, Prem Chandra Roy, and Subarna Shakya. "Temperature Prediction using Regression Model." Adv. Eng. ICT Converg. Proc 4 (2018): 161-170.
4. Singh, Siddharth, et al. "Weather forecasting using machine learning techniques." Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE). 2019.
5. Chen, Liuyi, et al. "Machine learning methods in weather and climate applications: A survey." *Applied Sciences* 13.21 (2023): 12019.
6. Jakaria, A. H. M., Md Mosharaf Hossain, and Mohammad Ashiqur Rahman. "Smart weather forecasting using machine learning: a case study in tennessee." *arXiv preprint arXiv:2008.10789* (2020)
7. Singh, Siddharth, et al. "Weather forecasting using machine learning techniques." *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*. 2019.
8. Naveen, L., and H. S. Mohan. "Atmospheric weather prediction using various machine learning techniques: a survey." *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2019..
9. Moosavi, Azam, Vishwas Rao, and Adrian Sandu. "Machine learning based algorithms for uncertainty quantification in numerical weather prediction models." *Journal of Computational Science* 50 (2021): 101295.
10. APAYDIN, Merve, et al. "Evaluation of air temperature with machine learning regression methods using Seoul City meteorological data." *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi* 28.5 (2022): 737-747.
11. Watson-Parris, Duncan. "Machine learning for weather and climate are worlds apart." *Philosophical Transactions of the Royal Society A* 379.2194 (2021): 20200098.
12. Karevan, Zahra, and Johan AK Suykens. "Transductive LSTM for time-series prediction: An application to weather forecasting." *Neural Networks* 125 (2020): 1-9.
13. Hewage, Pradeep, et al. "Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station." *Soft Computing* 24 (2020): 16453-16482.
14. Bochenek, Bogdan, and Zbigniew Ustrnul. "Machine learning in weather prediction and climate analyses—applications and perspectives." *Atmosphere* 13.2 (2022): 180.
15. Singh, Upma, et al. "A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments." *Energies* 14.16 (2021): 5196.
16. Liyew, Chalachew Muluken, and Haileyesus Amsaya Melese. "Machine learning techniques to predict daily rainfall amount." *Journal of Big Data* 8 (2021): 1-11.
17. Fister, Dusan, et al. "Accurate long-term air temperature prediction with Machine Learning models and data reduction techniques." *Applied Soft Computing* 136 (2023): 110118.
18. Cifuentes, Jenny, et al. "Air temperature forecasting using machine learning techniques: a review." *Energies* 13.16 (2020): 421