International Journal of Modern Computation, Information, and Communication Technology
2025;8(2):48-55.
ISSN: 2581-5954

Advances in Weather Forecasting Using Machine Learning Techniques

Ashish Kumar Dass[1*]

[1]*Department of Computer Science and Engineering, NIST Institute of Science and Technology, Berhampur, Odisha, India.*

[*]*Corresponding author: ashishkumardass@yahoo.co.in*

Abstract. Weather forecasting plays a crucial role in numerous sectors, ranging from agriculture and transportation to disaster management. Traditional meteorological models, while valuable, often face challenges in accurately predicting complex and dynamic weather patterns. This research paper explores the integration of machine learning (ML) techniques into weather forecasting to enhance predictive accuracy and reliability. The study begins by providing an overview of the limitations of conventional numerical weather prediction models and emphasizes the need for innovative approaches. It introduces a weather forecasting system based on machine learning, utilizing Decision Tree, Support Vector Machine, Random Forest, K-Nearest Neighbors, Gradient Boosting, Logistic Regression, and Naïve Bayes algorithms. The paper discusses the development and training of ML models using large datasets to capture intricate relationships among atmospheric variables. Among the evaluated models, Gradient Boosting achieves the highest predictive accuracy by effectively capturing nonlinear relationships and minimizing prediction errors. Performance evaluation demonstrates that integrating multiple machine learning techniques provides a stable, reliable, and scalable solution for short- to medium-term weather forecasting.

Keywords: Weather Forecasting, disaster management, Machine Learning, Atmospheric variables, Meteorological Data

INTRODUCTION

Weather forecasting has long been a critical aspect of modern society, influencing decisions ranging from agriculture and transportation to disaster preparedness. The ability to accurately predict atmospheric conditions is paramount for minimizing the impact of natural events and optimizing various human activities. Traditional numerical weather prediction models, rooted in physics-based simulations, have significantly advanced our understanding of atmospheric dynamics. However, the inherent complexity of atmospheric systems and the limitations of these models in capturing intricate patterns have spurred the exploration of alternative methodologies. In recent years, there has been a paradigm shift in the field of weather forecasting, driven by the integration of machine learning (ML) techniques. Machine learning, a subset of artificial intelligence, offers a data-driven approach that can uncover hidden patterns and relationships within vast and complex datasets. This has the potential to complement and, in some cases, surpass the capabilities of traditional models. As we delve into an era marked by unprecedented technological advancements, leveraging machine learning in weather forecasting emerges as a promising avenue to enhance predictive accuracy, extend lead times, and improve overall reliability. This research paper aims to explore and elucidate the advancements in weather forecasting facilitated by machine learning techniques. The integration of ML into meteorological practices opens new avenues for understanding and predicting atmospheric phenomena, providing valuable insights that can shape decision-making processes across various sectors. This introduction sets the stage for a comprehensive exploration of the methodologies, challenges, and practical applications associated with utilizing machine learning in weather forecasting, with the goal of contributing to the ongoing evolution of meteorological science and its societal impact.

RELATED WORKS

The foundation of modern meteorology relies on physics-based numerical weather prediction models, such as the Weather Research and Forecasting (WRF) model and the European Centre for Medium-Range Weather Forecasts (ECMWF) model [1]. These models simulate atmospheric processes using complex equations to predict weather patterns. However, their limitations in capturing fine-scale features and handling non-linear interactions

Received: 02.10.2025 Revised: 14.11.2025 Accepted: 28.11.2025

have spurred the exploration of alternative approaches [2]. The integration of machine learning techniques in meteorology has gained significant attention in recent years. The potential of support vector machines and neural networks in improving precipitation predictions [3]. Other studies explored the application of machine learning for short-term weather forecasting, emphasizing the advantages of data-driven approaches [4]. Feature selection plays a crucial role in developing accurate machine learning models for weather forecasting. The importance of selecting relevant atmospheric variables and the impact of feature engineering in capturing complex interactions within the data [5].

Ensemble methods, combining predictions from multiple models, have demonstrated superior performance in weather forecasting [6]. The study showcased the benefits of ensemble methods in probabilistic weather predictions, emphasizing their ability to quantify uncertainty and improve forecast reliability [7]. Despite the promising results, integrating machine learning into operational weather forecasting systems poses challenges [8]. Discusses issues such as model interpretability, data assimilation, and the need for real-time adaptability, highlighting the complexities of transitioning from research to practical implementation [9]. The rise of deep learning, particularly convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM), has shown promise in capturing spatial and temporal dependencies within weather data [10]. Notable studies demonstrate the efficacy of deep learning architectures in improving the accuracy of precipitation and temperature forecasts [11]. Numerous case studies validate the practical applicability of machine learning in weather forecasting. For instance, explores the use of machine learning in predicting extreme weather events, showcasing its potential in enhancing preparedness and response strategies [12]. While the integration of machine learning in weather forecasting has shown remarkable progress, there exist research gaps and opportunities for further exploration. Future directions include investigating the impact of climate change on machine learning-based predictions, refining data assimilation techniques, and addressing computational challenges for large-scale operational implementation [13].

PROPOSED SYSTEM

In response to the limitations of traditional numerical weather prediction models and the evolving landscape of machine learning in meteorology, the proposed system aims to leverage advanced machine learning techniques to enhance the accuracy and reliability of weather forecasting. The system outlined in this research introduces innovative methodologies and addresses specific challenges associated with integrating machine learning into operational forecasting systems. Figure 1 shows the machine learning life cycle.

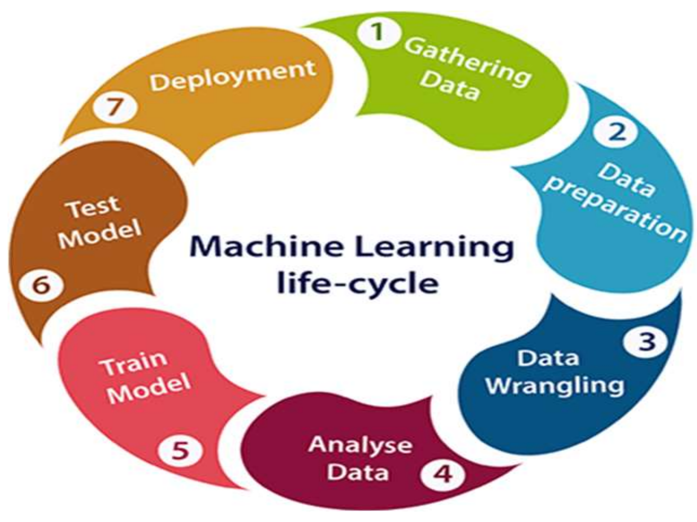

FIGURE 1. Machine Learning Life Cycle

The proposed system advocates for the development of hybrid machine learning models that combine the strengths of various algorithms. Ensemble methods, blending the predictions of multiple models, will be explored to harness the complementary strengths of different machine learning architectures. This approach seeks to mitigate the weaknesses of individual models, enhance predictive accuracy, and provide more robust forecasts.

The proposed system outlined in this research represents a holistic approach to integrating machine learning into weather forecasting. By combining various machine learning architectures, emphasizing real-time adaptability, ensuring model interpretability, and addressing scalability concerns, the system aims to advance the state-of-the-art in weather prediction and contribute to the ongoing evolution of meteorological practices.

Data Collection: Gather historical weather data from diverse sources, including ground-based weather stations, satellites, and other remote sensing instruments. Acquire datasets containing atmospheric variables such as temperature, humidity, wind speed, air pressure, and precipitation on both spatial and temporal scales.

Data Preprocessing: Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies. Normalize or standardize data to ensure uniformity and facilitate the convergence of machine learning algorithms.

Feature Selection and Engineering: Conduct a comprehensive analysis to identify relevant features crucial for accurate weather predictions. Explore the creation of derived features that capture complex interactions and patterns within the atmospheric data.

Model Selection: Evaluate and select appropriate machine learning algorithms based on the nature of the weather forecasting problem. Experiment with a range of models, including but not limited to neural networks, support vector machines, decision trees, and ensemble methods.

Training and Validation: Split the dataset into training and validation sets to train the machine learning models. Employ 10-fold cross-validation techniques to assess model generalization performance and mitigate over-fitting.

Model Evaluation: Assess the performance of machine learning models using appropriate evaluation metrics, considering factors like accuracy, precision, recall, and F1 score. Compare the results with traditional numerical weather prediction models to gauge improvements. The following chine learning models are employed.

- **Decision Tree:** Decision Tree [14] categorises weather data by using feature criteria to define circumstances such as sunny, wet, or stormy. It is interpretable, accommodates mixed data, but may overfit when presented with noisy datasets. Information Gain is used to decide the best feature to split.

$$IG(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \qquad (1)$$

- **Support Vector Machine (SVM):** SVM shows meteorological categories using an ideal hyperplane, forecasting severe occurrences or temperatures [15]. Efficient for high-dimensional datasets, accommodates nonlinear patterns using kernel methods, although requires meticulous parameter optimisation.

- **Random Forest:** Random Forest [16] integrates several decision trees to enhance precision in forecasting rainfall, temperature, or wind conditions. It mitigates overfitting, accommodates absent data, however, may exhibit diminished interpretability compared to an individual tree.

- **K-Nearest Neighbors (KNN):** KNN anticipates weather by analysing similarities to past data points, making it suitable for short-term predictions [17]. It is straightforward, non-parametric, however computationally intensive and susceptible to feature scaling and noisy data.

- **Gradient Boosting:** Gradient Boosting [18,21] constructs models to reduce forecast errors, effectively capturing nonlinear correlations in meteorological phenomena. Exceedingly precise at forecasting precipitation or temperature, however susceptible to overfitting if inadequately calibrated.

- **Logistic Regression:** Logistic Regression calculates the odds of binary or categorical meteorological occurrences, such as the likelihood of rain or storms [19]. It is straightforward and comprehensible, although less efficacious for intricate nonlinear meteorological correlations. The Sigmoid function maps feature to probability of event.

$$P(y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}} \tag{2}$$

- **Naïve Bayes:** Naïve Bayes [20] forecasts meteorological conditions by probabilistic reasoning, predicated on the premise of feature independence. Rapid and effective, appropriate for limited datasets; nonetheless, interrelations across meteorological variables may reduce accuracy.

Deployment: Implement methods and deploy it to enhance the interpretability of machine learning models, ensuring that forecasters can comprehend and trust the predictions. The system employs many machine learning techniques on the pre-processed data.

RESULTS AND DISCUSSION

The success of machine learning models in weather forecasting heavily relies on the quality and diversity of the datasets used for training, validation, and testing. The dataset selected for this research aims to encompass a comprehensive range of atmospheric variables, temporal scales, and geographical locations to ensure the robustness and generalization of the developed machine learning models. Integrate datasets that document historical weather events, including extreme events such as hurricanes, tornadoes, and heat waves. This information is crucial for evaluating the performance of machine learning models in predicting and responding to high-impact weather conditions. Explore open-access meteorological databases provided by organizations like the National Oceanic and Atmospheric Administration (NOAA) and the European Space Agency (ESA). These databases offer a wealth of meteorological data covering various spatiotemporal scales. The selection of this diverse and extensive dataset aims to provide a rich and representative set of information for training and evaluating machine learning models in weather forecasting. The combination of ground-based observations, satellite imagery, reanalysis data, and real-time information ensures that the developed models can handle the complexity and variability inherent in atmospheric conditions. Figure 2 to Figure 6 show the various outputs and their corresponding codes are also given below:

```
#graph himidity vs temparture
plt.scatter(np.log10(dataset[' _hum']),dataset[' _tempm'])
plt.title('humidity vs temprature')
plt.xlabel("------------humidity-------------")
plt.ylabel("------------temprature-----------")
plt.show()
```

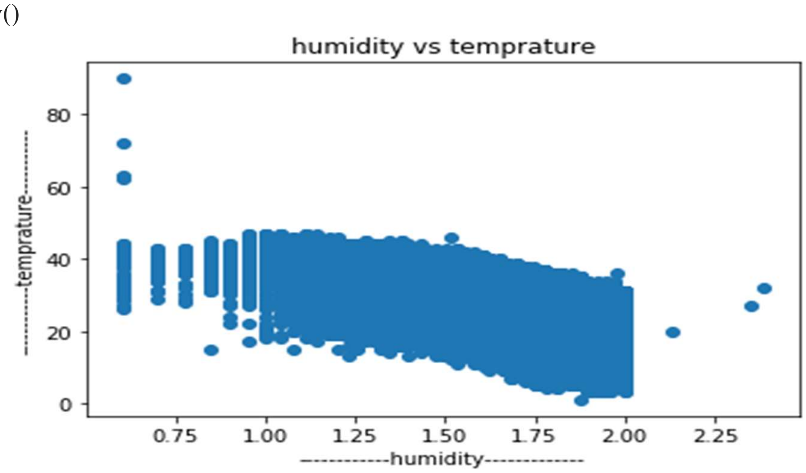

FIGURE 2. Humidity vs. Temperature

```
#histogram of data how they looks like on graphical represantation
plt.hist(dataset[' _tempm'],facecolor='red',edgecolor='blue',bins=50,range=(5,35))
plt.title("temperature histogram")
plt.ylabel("------no. of occurrence temperature values----")
plt.show()
```

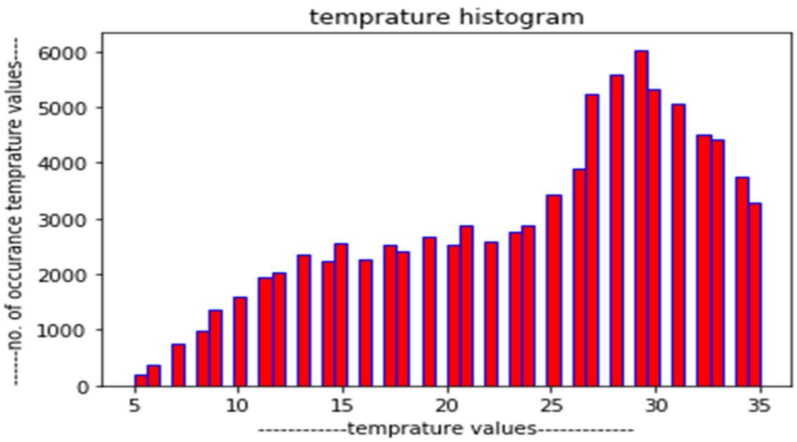

FIGURE 3. Temperature histogram

```
#graph dewpoint vs temparture
plt.scatter(dataset[' _dewptm'],dataset[' _tempm'])
plt.title(' dewpoint vs temprature')
plt.xlabel("----------- dewpoint-------------")
plt.ylabel("------------temprature-----------")
plt.show()
```

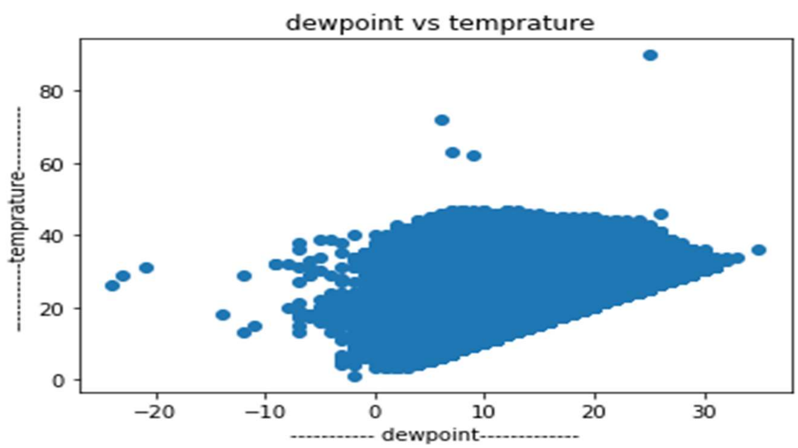

FIGURE 4. Dewpoint vs. Temperature

```
plt.title(" dewpoint histogram")
plt.xlabel("------------dewpoint values-------------")
plt.ylabel("------no. of occurance dewpoint values----")
plt.show()
```

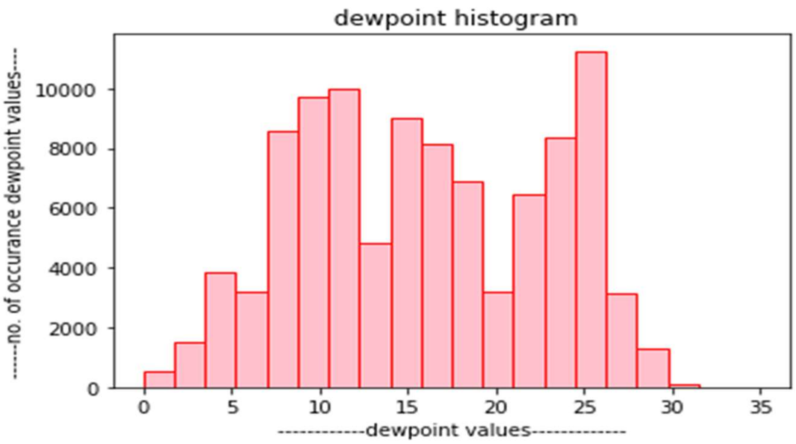

FIGURE 5. Dewpoint Histogram

```
training & testing for temprature prediction @ data size 0.2:-
y_prediction=model.predict(X_test)
score2=r2_score(y_test,y_prediction)
print("Temprature prediction Accuracy @test_size=0.2= ",score2*100)
dataset.shape
Temprature prediction Accuracy @test_size=0.2=  88.97643526246964
(99396, 10)
#histogram of data how they looks like on graphical represantation
plt.hist(y_prediction,facecolor='red',edgecolor='blue',bins=10,range=(5,35))
plt.title("predicted temprature histogram @test_size=0.2")
plt.xlabel("------Predicted temprature values-------------")
plt.ylabel("------no. of occurance temprature values----")
plt.show()
```

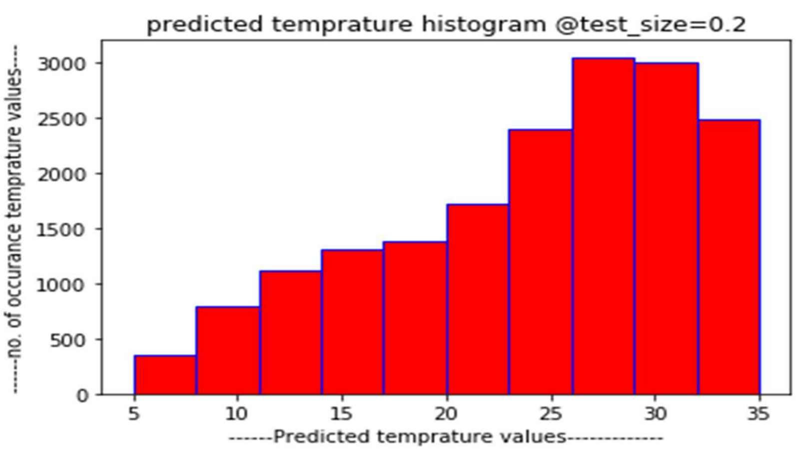

FIGURE 6. Predicted temperature histogram

Accuracy is a metric that measures how often a machine learning model correctly predicts the outcome. It is the number of correct predictions divided by the total number of predictions across all classes. In this experiment we have gone through Decision Tree, Support vector machine, Random Forest, KNN algorithm, gradient boosting, Logistic regression and Naïve Bayes methods and the results are shown below.

TABLE 1. Accuracy of Different Models

Sl Num	Model Name	Accuracy
1	Decision Tree	71.52
2	Support Vector Machine	60.55
3	Random Forest	78.42
4	KNN	75.78
5	Gradient boosting	80.87
6	Logistic regression	75.15
7	Naïve Bayes	72.88

CONCLUSION

This research endeavors to contribute to the ongoing evolution of weather forecasting by investigating and harnessing the potential of machine learning techniques. The exploration of advancements in machine learning for weather forecasting has revealed promising avenues for improving predictive accuracy, extending lead times, and enhancing overall reliability. The proposed system, integrating hybrid machine learning models, deep learning architectures, real-time data assimilation, and continuous model updating, represents a comprehensive approach to address the challenges faced by traditional numerical weather prediction models. Out of the entire model it has found that Gradient boosting method gives highest accuracy in comparison to other models.

REFERENCES

[1]. G. Camps-Valls, A. García-Martín, and M. Campos-Taberner, 2018, “Advances in Remote Sensing Image Retrieval,” IEEE Transactions on Geoscience and Remote Sensing, 56(9), pp. 5370–5383.

[2]. D.J. Gagne, G. Camps-Valls, and C. Rudin, 2019, “On the Limitations of Representing Functionality in Learning Weather Dependent Models,” Journal of Climate, 32(23), pp. 8239–8253.

[3]. D.J. Lary, L.A. Remer, J. Macri, R.G. Kleidman, and R.C. Levy “A New Technique for the Simultaneous Retrieval of Aerosol and Cloud Optical Thickness and Effective Radius Using Airborne Measurements,” Atmospheric Measurement Techniques, 9(8), pp. 3245–3262.

[4]. Ashish Kumar Dass, & Manjushree Nayak, 2023, “Comparative Analysis of Machine Learning and Neural Network Models for Wine Quality Prediction,” Advancement of Computer Technology and Its Applications, 6(3), 34–51.

[5]. A.E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005, “Using Bayesian Model Averaging to Calibrate Forecast Ensembles,” Monthly Weather Review, 133(5), pp. 1155–1174.

[6]. X. Shi, Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, and W.C. Woo, 2015, “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting,” In Advances in Neural Information Processing Systems, pp. 802–810.

[7]. Ashish Kumar Dass, “Comparison of Heart Disease Prediction Using Different Machine Learning Algorithms,” 05 February 2023, PREPRINT (Version 1) available at Research Square.

[8]. B. Casati, M. Scheuerer, and F. Pappenberger, 2020, “Probabilistic Wind Speed Forecasts for the Power Industry Using Ensemble Model Output Statistics,” Applied Energy, 260, Article. 114163.

[9]. P.D. Dueben, and P. Bauer, 2018, “Challenges and Design Choices for Global Weather and Climate Models Based on Machine Learning,” Geoscientific Model Development, vol. 11, no. 10, pp. 3999–4009.

[10]. M. Fan, O. Imran, A. Singh, and S. A. Ajila, 2022, "Using CNN-LSTM Model for Weather Forecasting,” IEEE International Conference on Big Data (Big Data), pp. 4120-4125.

[11]. G. Zenkner, and S. Navarro-Martinez, 2023, “A flexible and lightweight deep learning weather forecasting model,” Applied Intelligence, 53(21), pp. 24991-25002.

[12]. P. Das, P. Parmar, S. Sahoo, A. Saluja and S. Pande, 2024, "An Intelligent Regression Approach for Weather Forecasting System Using Machine Learning," 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), pp. 1-6,

[13]. S. Kaur, A. P. Singh, A. Pandey and H. Chauhan, 2023, "Real-Time Weather Forecasting Using Machine Learning," Seventh International Conference on Image Information Processing (ICIIP), pp. 248-253.

[14]. A. Noeman, D. Handayani, and A. Hiswara, “Decision Tree-Based Weather Prediction. PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic,” 10(1), pp. 67-78.

[15]. A.V. Wulandari, N.J. Trilaksono, and M. Ryan, 2024, “Improving Short-Term Weather Forecasting using

Support Vector Machine Method in North Barito," Jurnal Meteorologi dan Geofisika, 25(2), pp. 113-121.

[16]. A. Mathew, and J. Mathew, 2022, "Weather forecasting using the random forest algorithm analysis," In Proceedings of the National Conference on Emerging Computer Applications, 4(1), pp. 1-15.

[17]. T. Benil, P.K. Kumar, R. Bharathi, "Efficient data pruning using optimal KNN for weather forecasting in cloud computing," International Journal of Global Warming, 30(2), pp. 137-151.

[18]. S. Babu Nuthalapati, and A. Nuthalapati, 2024, "Accurate weather forecasting with dominant gradient boosting using machine learning," International Journal of Science Research Archive, 12(2), pp. 408-422.

[19]. K. Shen, 2024, "Applying Machine Learning Technology for Weather Forecasting: A Case Study of the Logistic Regression Model," In Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition, pp. 572-576.

[20]. M. Biswas, T. Dhoom, and S. Barua, 2018, "Weather forecast prediction: an integrated approach for analysing and measuring weather data," International Journal of Computer Applications, 182(34), pp. 20-24.

[21]. C. R. G. A. A'syifa, B. A. Pramudita, Yudiansyah, B. P. A. Rohman, and D. P. Setiawan, 2025, "Comparing Predictive Capabilities of Machine Learning Models in Weather Forecasting," International Conference on Data Science and Its Applications, pp. 973-978.