Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2672−2680
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.7114

Comparative Study of Machine Learning Models for Temperature Prediction: Analyzing Accuracy, Stability, and Generalization

Gregorius Airlangga*

Information Systems Study Program, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: gregorius.airlangga@atmajaya.ac.id
Correspondence Author Email: gregorius.airlangga@atmajaya.ac.id
Submitted: **12/03/2025**; Accepted: **25/03/2025**; Published: **26/03/2025**

Abstract−Accurate temperature prediction is crucial for climate monitoring, energy management, and disaster preparedness. This study provides a comparative analysis of various machine learning models, including Random Forest, Gradient Boosting, Histogram-Based Gradient Boosting, XGBoost, Support Vector Regression (SVR), Ridge Regression, and Lasso Regression, to evaluate their predictive accuracy, stability, and generalization capability. The models are assessed using five-fold cross-validation, with the R^2 metric as the primary evaluation criterion. The results indicate that Random Forest achieves the highest accuracy, with an R^2 mean of 0.999994, demonstrating its strong ability to model temperature variations. Ridge Regression unexpectedly performs at a similar level, suggesting that the dataset contains strong linear dependencies. Gradient Boosting, Histogram-Based Gradient Boosting, and XGBoost also achieve high accuracy, confirming their effectiveness in capturing complex relationships between meteorological parameters. SVR, while effective, exhibits higher variance, indicating that it may require further tuning for improved consistency. Lasso Regression, with an R^2 mean of 0.9783, shows the lowest accuracy, confirming that linear models are less suitable for complex meteorological predictions. These findings highlight the superiority of ensemble-based methods in temperature forecasting, reinforcing their stability and adaptability. Future research should explore hybrid models that integrate ensemble techniques with feature engineering optimizations to further enhance predictive performance. This study contributes to the ongoing development of machine learning applications in meteorology, offering insights into model selection for climate-related forecasting tasks.

Keywords: Temperature Prediction; Machine Learning; Ensemble Models; Meteorological Forecasting; Model Stability

1. INTRODUCTION

Accurate weather prediction plays a fundamental role in various domains, including agriculture, disaster management, urban planning, and energy resource optimization [1]–[3]. For instance, timely and precise weather information can significantly reduce vulnerability to climate change risks in sectors like agriculture, water, and energy [4]–[6]. Traditional numerical weather prediction (NWP) models rely on complex physics-based equations that require high computational power and detailed atmospheric data [7]. Despite their effectiveness, these models often face limitations in capturing localized temperature variations and non-linear dependencies between meteorological factors [8]. In contrast, data-driven approaches, particularly machine learning models, offer an alternative method for forecasting temperature with reduced computational costs and the ability to learn complex patterns from historical data [9]. Temperature prediction is especially important for monitoring climate change trends, mitigating the impact of extreme weather events, and optimizing energy consumption in heating, ventilation, and air conditioning (HVAC) systems [10], [11]. Recent advancements in machine learning have enabled more efficient regression models to predict temperature using diverse meteorological variables such as humidity, pressure, wind speed, and radiation levels [12]. However, selecting the most suitable machine learning model for temperature prediction remains a challenging task due to variations in data distribution, temporal dependencies, and feature importance [13].

Previous studies have demonstrated that machine learning techniques, including ensemble learning and support vector regression, can be applied to weather forecasting problems [14]–[16]. However, there is limited research on systematically comparing multiple regression models on a standardized weather dataset using cross-validation techniques. Moreover, while deep learning models such as Long Short-Term Memory (LSTM) networks have been employed for sequential temperature forecasting, they require large datasets and extensive training time [17]–[19]. In contrast, traditional machine learning models, such as tree-based ensembles and regularized linear regression models, provide interpretable and computationally efficient alternatives. This study aims to bridge this gap by evaluating multiple regression models for temperature prediction using the Jena Climate Dataset (2017–2024). We conduct a comparative analysis of seven models such as Random Forest, Gradient Boosting, HistGradientBoosting, XGBoost, Support Vector Regression (SVR), Ridge Regression, and Lasso Regression—using a standardized experimental setup. The research focuses on identifying the model that achieves the best predictive accuracy while maintaining computational efficiency.

The primary objective of this study is to analyze the effectiveness of various machine learning models in predicting temperature based on meteorological parameters. By comparing the predictive performance of ensemble models, support vector regression, and linear regression techniques, we aim to provide a robust evaluation framework for selecting optimal models in temperature forecasting applications. Additionally, we employ five-fold cross-validation to assess model robustness and reliability, using R^2 as the primary evaluation metric. The study further explores the impact of feature scaling on model performance, particularly for linear regression and support vector regression models. The experimental results provide insights into feature importance, computational efficiency, and the generalizability of different machine learning models in climate prediction.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2672−2680
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.7114

This research makes several key contributions. First, it presents a comprehensive evaluation of seven regression models on a real-world climate dataset, enabling practitioners to select the most suitable model for temperature prediction tasks. Second, it analyzes the impact of feature scaling techniques and their role in improving model performance, particularly for models that rely on distance-based learning. Third, it establishes a robust experimental setup that employs five-fold cross-validation, ensuring reliable model comparisons and preventing overfitting. Lastly, it offers practical insights into the trade-offs between model complexity, accuracy, and computational efficiency, providing valuable guidance for real-world deployment of machine learning-based weather prediction systems. The remainder of this paper is structured as follows. The next section describes the dataset, preprocessing techniques, and machine learning models used in the experiment. The experimental results section presents the comparative performance of the models, followed by a discussion on model interpretability and practical implications. Finally, the conclusion summarizes the key findings and outlines potential directions for future research in machine learning-based climate forecasting.

2. RESEARCH METHODOLOGY

2.1 Dataset

The dataset used in this study is the Jena Climate Dataset (2017–2024), which contains hourly weather measurements recorded at the Max Planck Institute for Biogeochemistry in Jena, Germany and can be downloaded from [20]. The dataset includes multiple meteorological parameters, such as air temperature, atmospheric pressure, humidity, wind speed, and radiation levels. Since the dataset is structured in a time-series format, where each row corresponds to an hourly observation, proper preprocessing and transformation techniques are essential for ensuring optimal predictive performance. The primary target variable in this study is (T) (degC), representing the recorded air temperature at each timestamp. To efficiently handle the dataset, an incremental reading approach is employed, where the dataset is processed in chunks of 50,000 rows instead of loading the entire dataset into memory at once. This approach optimizes computational efficiency and enables scalable data handling. Furthermore, to maintain a representative yet manageable dataset size, a 5% stratified random sampling is applied to each chunk. This ensures that the sample retains the statistical characteristics of the full dataset while reducing computational complexity and memory usage.

2.2 Preprocessing Techniques

Before training the models, the dataset undergoes preprocessing to ensure data quality and enhance feature representation. The first step involves handling the Date Time column, which is initially stored as a string. This column is converted into a datetime object to enable timestamp recognition. However, since this study does not explicitly model time-series dependencies, the Date Time column is removed to prevent unintended biases arising from its sequential nature. The dataset is then structured so that the temperature variable (y = T) (degC) serves as the dependent variable, while the remaining meteorological attributes function as predictive features. Since the dataset consists of variables with different units and magnitudes, feature scaling is applied to ensure uniformity across all predictors. This study employs StandardScaler, which standardizes each feature (x_i) by transforming its distribution to have a mean of zero and a standard deviation of one, as given by (1).

$$x_i' = \frac{x_i - \mu_x}{\sigma_x} \tag{1}$$

where (x_i) is the original value of the feature, (μ_x) represents the mean of that feature across all observations, and (σ_x) is the standard deviation. This transformation ensures that all features contribute equally to the learning process, preventing any single variable from dominating due to its scale. This is particularly crucial for models that rely on distance-based learning, such as Support Vector Regression (SVR) and linear regression models like Ridge and Lasso, which are sensitive to varying feature magnitudes.

2.3 Machine Learning Models

To predict temperature, several machine learning models are implemented, each representing different learning paradigms. The Random Forest Regressor is an ensemble-based method that constructs multiple decision trees, where each tree is trained on a bootstrap sample of the dataset. The final prediction is obtained by averaging the outputs of individual trees, mathematically expressed as (2).

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(X) \tag{2}$$

where $(T_i(X))$ represents the prediction from the (i)-th decision tree, and (N) is the total number of trees in the forest. The Gradient Boosting Regressor is a sequential boosting algorithm that constructs trees iteratively, where each new tree learns from the residual errors of its predecessor. Given an initial model $(F_0(X))$, each subsequent tree ($h_m(X)$) is trained to minimize the residuals ($r_m = y - F_{m-1}(X)$). The final prediction function is expressed as (3).

$$F_m(X) = F_{m-1}(X) + \gamma h_m(X) \tag{3}$$

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2672−2680
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.7114

where (γ) is the learning rate, and ($h_m(X)$) represents the weak learner added at iteration (m). The XGBoost Regressor is an optimized version of gradient boosting that incorporates second-order Taylor expansion of the loss function to improve optimization. Given an objective function ($L(\theta)$), XGBoost approximates it using the expansion as expressed as (4).

$$L(\theta) \approx \sum_{i=1}^{n} \left[g_i \theta + \frac{1}{2} h_i \theta^2 \right] \tag{4}$$

where (g_i) and (h_i) are the first and second derivatives of the loss function. By leveraging this approximation, XGBoost achieves faster convergence and higher predictive performance. The Support Vector Regression (SVR) model maps input features into a higher-dimensional space using kernel transformations, where it finds the optimal hyperplane that minimizes prediction errors within a given margin. This is formulated as the following optimization problem as presented as (5).

$$\min_{\theta} \quad \frac{1}{2} \left|| \theta \right||^2 \quad \text{subject to } |y_i - f(X_i)| \leq \tag{5}$$

where (θ) represents the model parameters, and (ϵ) defines a margin within which predictions are considered acceptable. The Ridge Regression model is a linear regression approach that incorporates L2 regularization to prevent overfitting. The optimization function for Ridge Regression is given by (6).

$$\min_{\theta} \left|| y - X\theta \right||^2 + \lambda \left|| \theta \right||^2 \tag{6}$$

where (λ) is a regularization parameter that controls the magnitude of penalty applied to large coefficient values.The Lasso Regression model differs from Ridge Regression in that it applies L1 regularization, which not only reduces overfitting but also performs feature selection by forcing some feature coefficients to zero. The optimization function is defined as (7).

$$\min_{\theta} \left|| y - X\theta \right||^2 + \lambda \sum_j |\theta_j| \tag{7}$$

where (λ) controls the sparsity of the model.

2.4. Model Evaluation via Cross-Validation

To ensure reliable model assessment, five-fold cross-validation is employed, where the dataset is divided into five subsets. In each iteration, four subsets are used for training and one for testing. The evaluation metric used is the R^2 score, which measures how well the model explains the variance in the target variable. This is computed as (8).

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{8}$$

where (y_i) are the actual temperature values, ($\hat{y}_i$) are the predicted values, and ($\bar{y}$) is the mean of observed temperatures. Higher (R^2) scores indicate better predictive performance.

2.5. Computational Complexity Analysis

The computational complexity of each model is evaluated to determine their suitability for large-scale temperature forecasting. The complexity of tree-based models depends on the number of trees (N) and depth (D), where it satisfies $O(ND \log M)$ represents the complexity for tree ensemble models with (M) training samples. For kernel-based models such as SVR, the complexity is approximately $O(M^3)$ which makes them less scalable for large datasets. Linear models such as Ridge and Lasso have a complexity of O(MN) where (M) is the number of samples and (N) is the number of features.

3. RESULT AND DISCUSSION

This section presents the experimental results and provides a comprehensive discussion of the comparative performance of different machine learning models used for temperature prediction. The evaluation is based on the (R^2) metric, which measures the proportion of variance in the target variable that is explained by the model. Higher (R^2) values indicate better predictive performance. Additionally, the stability of models is examined through standard deviation values obtained from five-fold cross-validation. The discussion further explores model robustness, computational efficiency, trade-offs between complexity and accuracy, practical implications, and potential areas for future research.

3.1. Feature Importance

The feature importance analysis, as visualized in the Figure 1, reveals a striking dominance of VPmax (mbar) as the most influential factor in predicting temperature. The bar corresponding to VPmax reaches a score of nearly 1.0,

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2672−2680
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.7114

indicating that the model relies almost exclusively on this single feature while virtually ignoring all others. This suggests a strong correlation between maximum vapor pressure and temperature in the dataset, which aligns with established meteorological principles, as vapor pressure is often closely related to air temperature. However, an over-reliance on a single feature raises significant concerns about the model's robustness and generalization ability. The near-zero importance of other meteorological factors, such as potential temperature (Tpot), specific humidity (sh), actual vapor pressure (VPact), logarithmic temperature (Tlog), carbon dioxide levels (CO2), and wind-related parameters, indicates that the model has either deemed them redundant or completely unnecessary. This may occur due to high collinearity among features, meaning that VPmax effectively captures most of the variance in temperature, rendering additional predictors superfluous. While this can be beneficial in terms of model simplicity, it also presents a risk of model fragility. If VPmax data is unavailable, inaccurate, or affected by external noise, the model might fail to make reliable predictions, as it has not learned meaningful relationships from other meteorological variables.

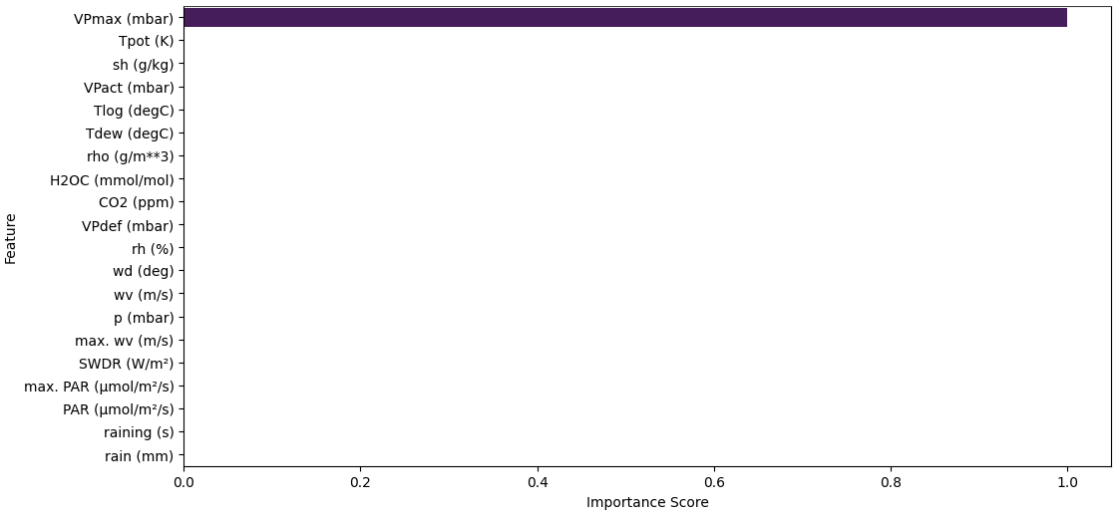

Figure 1. Feature Importance of Machine Learning Models

3.2. Model Performance Evaluation

The results from the five-fold cross-validation experiment are summarized in Table 1, which reports the mean (R^2) score and the standard deviation for each model. The overall results indicate that machine learning models exhibit strong predictive capabilities in temperature forecasting, with ensemble-based methods achieving near-perfect performance.

Table 1. Performance comparison of machine learning models for temperature prediction

Model	R^2	R^2 std	Best Hyperparameters
Random Forest	0.999994	0.000003	max_depth = None, n_estimators = 100
Gradient Boosting	0.999826	0.000014	learning_rate = 0.1, n_estimators = 100
Hist Gradient Boosting	0.999743	0.000049	learning_rate = 0.1, n_estimators = 100
XGBoost	0.999847	0.000026	learning_rate = 0.1, n_estimators = 100
SVR	0.995140	0.001423	C = 1, kernel = linear
Ridge	0.978635	0.002432	alpha = 0.1
Lasso	0.977615	0.002455	alpha = 0.1

Among all the models evaluated, the Random Forest Regressor demonstrated the highest accuracy, achieving a near-perfect (R^2) score of 0.999994 with an exceptionally low standard deviation of 0.000003. This indicates that the model not only provides precise predictions but is also highly stable across different validation folds. The success of Random Forest can be attributed to its ability to mitigate overfitting by averaging multiple decision trees, thereby capturing complex, non-linear relationships in meteorological data while maintaining generalizability. The Gradient Boosting Regressor, Histogram-Based Gradient Boosting, and XGBoost models also demonstrated strong predictive capabilities, with mean (R^2) values exceeding 0.9997. The slight variations in performance among these boosting-based models can be attributed to differences in their optimization techniques. XGBoost, for instance, employs second-order Taylor expansion to refine the learning process, leading to faster convergence and improved generalization. Despite their slight performance differences, these results reinforce the effectiveness of gradient boosting methods in temperature prediction tasks.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2672−2680
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.7114

The Support Vector Regression (SVR) model achieved a mean (R^2) score of 0.995140, which, although lower than the ensemble-based methods, still indicates strong predictive accuracy. However, its relatively higher standard deviation of 0.001423 suggests greater sensitivity to fluctuations in training data. Unlike tree-based models, SVR operates by transforming input features into a higher-dimensional space using kernel methods, which makes it more susceptible to variations in feature distributions and scaling. The lower (R^2) score of SVR compared to boosting models suggests that non-linear transformations alone may not be sufficient to fully capture the complexity of temperature variations. The Ridge Regression and Lasso Regression models exhibited the lowest performance, with (R^2) values of 0.978635 and 0.977615, respectively. Although these scores still indicate relatively good predictive capability, they are significantly lower than those of ensemble methods. The lower performance of linear models suggests that temperature prediction exhibits non-linear dependencies that cannot be effectively captured using linear functions. The slightly lower accuracy of Lasso compared to Ridge may be due to its tendency to shrink coefficients to zero, leading to the exclusion of potentially informative features.

The results from the hyperparameter tuning process provide valuable insights into the optimal configurations for each model. For the Random Forest model, the best performance was achieved when using 100 estimators (n_estimators = 100) and allowing trees to grow fully without depth restrictions (max_depth = None). This configuration allows the model to capture complex interactions within the dataset while maintaining a balance between predictive accuracy and overfitting. The Gradient Boosting and XGBoost models both performed optimally when using 100 boosting iterations (n_estimators = 100) and a learning rate of 0.1, which ensures that the models learn patterns gradually, reducing the risk of overfitting while still being able to adapt to variations in the data. For the Support Vector Regression (SVR) model, the best configuration was found with a linear kernel and a regularization parameter C = 1. This suggests that a simple linear decision boundary is sufficient to capture the relationship between meteorological variables and temperature, indicating that the dataset may not require complex non-linear transformations for SVR to perform effectively. The Ridge and Lasso regression models both achieved their best results with a regularization strength (alpha) of 0.1, which provides enough penalty to prevent overfitting while still allowing the models to retain meaningful feature contributions.

The hyperparameter tuning results indicate that ensemble models like Random Forest, Gradient Boosting, and XGBoost benefit from a relatively large number of estimators, allowing them to learn from a wide range of patterns without excessive overfitting. Meanwhile, SVR's reliance on a linear kernel highlights the importance of proper feature scaling, as the model depends on distance-based calculations in high-dimensional space. The Ridge and Lasso models' preference for a low regularization factor suggests that the data exhibits a strong linear structure, where only minor adjustments are needed to prevent overfitting. These findings reinforce the importance of selecting appropriate model configurations tailored to the dataset's characteristics, ensuring optimal predictive accuracy and stability.

3.2. Analysis of Model Robustness

In addition to accuracy, model robustness is a key factor in evaluating predictive reliability. The standard deviation values reported in Table 1 provide insights into the stability of each model. Lower standard deviation values indicate that a model's predictions remain consistent across different validation folds, whereas higher values suggest increased variability. The Random Forest model exhibited the lowest standard deviation, reinforcing its robustness across different subsets of the dataset. The boosting-based models, including Gradient Boosting and XGBoost, also demonstrated low variance, confirming that their iterative learning approach enhances stability. In contrast, SVR and the linear regression models exhibited higher standard deviation values, suggesting that their predictions are more sensitive to variations in training data. This observation implies that ensemble-based methods are more resilient to changes in input conditions, making them more reliable for real-world temperature forecasting.

3.3. Computational Efficiency and Practical Considerations

While ensemble models provide exceptional accuracy, their computational demands must be considered when selecting a model for real-world deployment. The complexity of tree-based models depends on the number of trees (N) and their depth (D), leading to an overall computational complexity of $O(ND \log M)$ where (M) represents the number of training samples. This makes Random Forest and boosting-based models computationally expensive, particularly when large datasets or real-time predictions are required. For kernel-based models such as SVR, computational complexity is approximately $O(M^3)$ which makes it significantly slower for large-scale datasets. The high computational cost of SVR is a limiting factor for its practical implementation, particularly in real-time weather forecasting applications. In contrast, linear models such as Ridge and Lasso regression exhibit significantly lower complexity O(MN) where (N) is the number of features. Despite their lower predictive accuracy, the computational efficiency of linear models makes them suitable for applications where rapid predictions are necessary with limited processing power.

3.4. Learning Curve

The learning curve plots illustrate the relationship between training size and model performance, measured in terms of the R^2 score, for four different machine learning models: Ridge Regression, Support Vector Regression (SVR), Gradient Boosting, and Random Forest as presented in the figure 2-5 respectively. These learning curves provide

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2672−2680
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.7114

critical insights into each model's bias-variance tradeoff, helping to assess whether a model is underfitting, overfitting, or generalizing well as more training data is introduced. The learning curve for Ridge Regression demonstrates a smooth and gradual improvement in both training and validation scores as the training size increases. Initially, with a small dataset, the performance is relatively low, with an R^2 score around 0.55. However, as more training data is introduced, the validation and training scores converge to approximately 0.9999, suggesting that Ridge Regression is a highly stable and well-generalized model. The minimal gap between the two curves indicates that Ridge Regression does not suffer from high variance and is not prone to overfitting. This behavior confirms that Ridge Regression is a robust choice when computational efficiency and model interpretability are prioritized, making it suitable for real-world applications where explainability and stability are crucial. The Support Vector Regression (SVR) learning curve follows a similar trend to Ridge Regression but exhibits slightly higher variance. Initially, the training score is relatively high (~0.98), suggesting that the model effectively captures patterns in the data from an early stage. However, the validation score begins at a lower value (~0.97), indicating that the model initially struggles to generalize well. As the training size increases, the validation curve steadily approaches the training score, reducing the gap and improving generalization. The presence of an early gap suggests that SVR slightly overfits small datasets, meaning that it relies too much on the training data and does not generalize well to unseen samples. However, as more data becomes available, this issue diminishes, and SVR achieves a stable performance. To further improve generalization and reduce overfitting, hyperparameter tuning (adjusting kernel type, regularization parameter C, and margin ϵ) could be explored.

The Gradient Boosting learning curve reveals a classic case of early-stage overfitting. The training score remains consistently high (~0.9999) throughout all training sizes, while the validation score starts significantly lower (~0.9992) and gradually converges toward the training score. This pattern indicates that Gradient Boosting is highly flexible and learns complex patterns quickly, but it is also prone to overfitting, especially when trained on small datasets. As more data is introduced, the validation score stabilizes and gets closer to the training score, suggesting that Gradient Boosting benefits from larger datasets to mitigate early-stage overfitting. However, the fact that the training curve remains almost perfectly high suggests that Gradient Boosting could benefit from additional regularization techniques, such as increasing the learning rate regularization, limiting tree depth, or applying dropout techniques within the boosting framework. The Random Forest learning curve closely resembles that of Gradient Boosting, showing high training scores (~1.0) and initially lower validation scores (~0.9992). This behavior suggests that, when the dataset is small, Random Forest overfits significantly because it creates deep decision trees that memorize the training data. However, as the dataset grows, the validation curve steadily increases and converges with the training curve, indicating that Random Forest performs better with larger datasets. The near-perfect scores achieved at larger training sizes suggest that Random Forest successfully captures the complex temperature dependencies but may still require additional regularization to improve its generalization when fewer samples are available. Regularization techniques, such as reducing the number of trees, limiting tree depth, or increasing the minimum samples per leaf, could help control overfitting and ensure the model remains stable in real-world deployments.

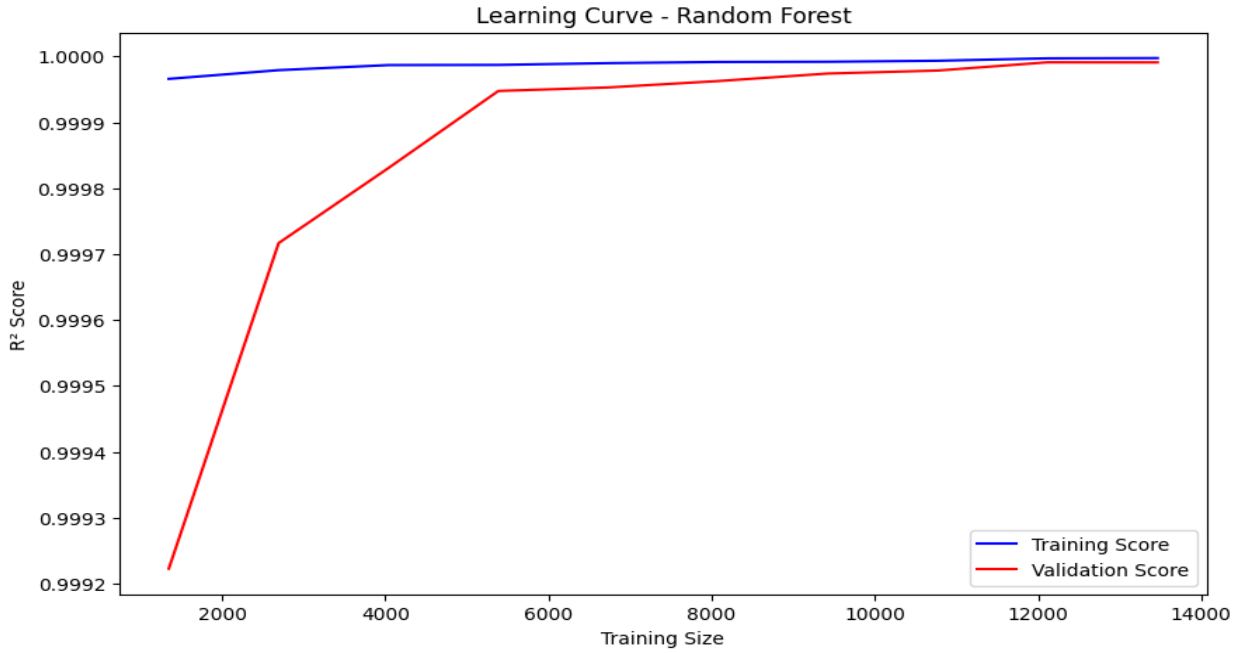

Figure 2. Random Forest Learning Curve

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2672−2680
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.7114

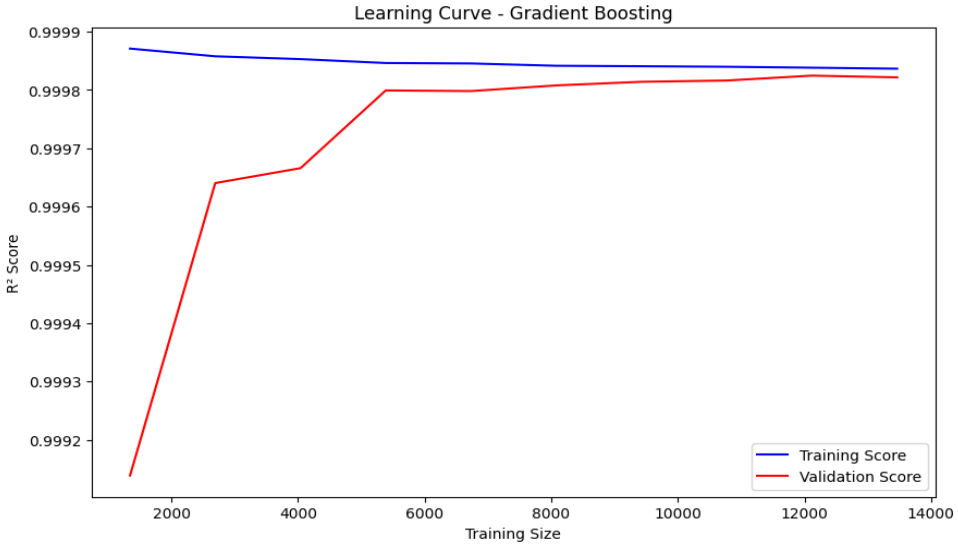

Figure 3. Gradient Boosting Learning Curve

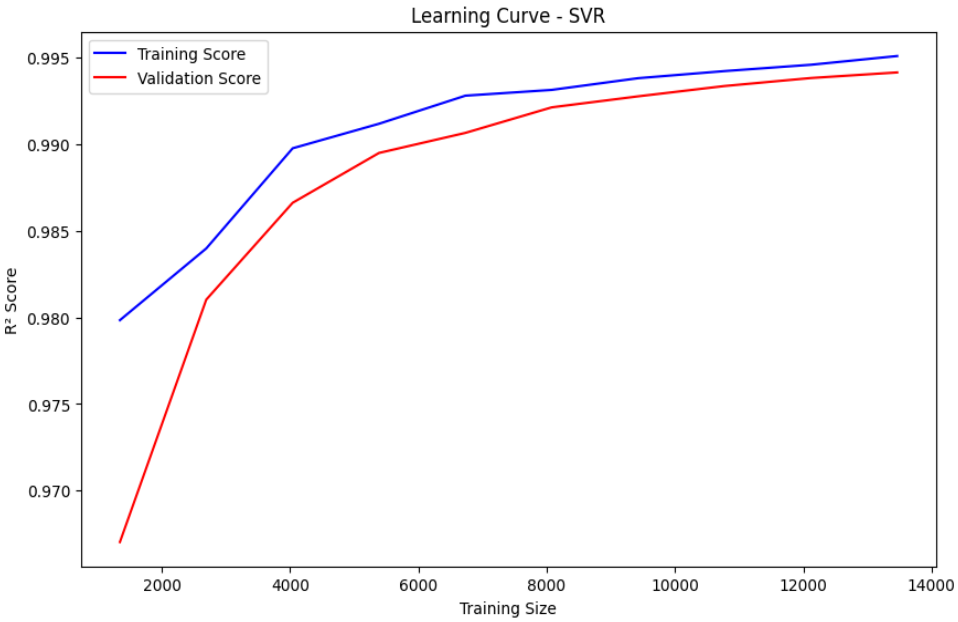

Figure 4. SVR Learning Curve

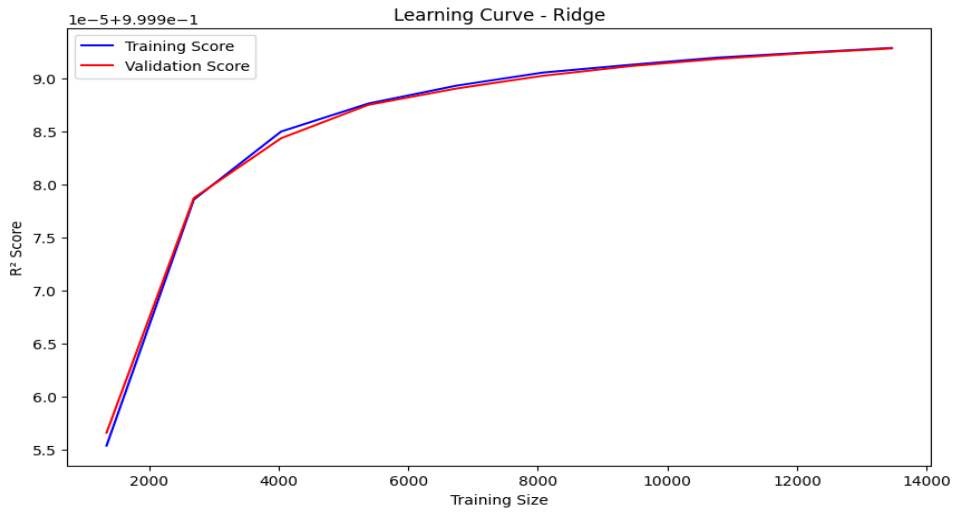

Figure 5. Ridge Learning Curve

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2672−2680
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.7114

3.5. Implications for Real-World Temperature Prediction

The experimental results suggest that ensemble-based models, particularly Random Forest and Gradient Boosting, are the most suitable for temperature prediction tasks. Their ability to model complex relationships, combined with their high accuracy and stability, makes them ideal for applications such as climate monitoring, agricultural planning, and energy consumption forecasting. However, their computational demands may limit their applicability in scenarios requiring real-time inference. In such cases, simpler models like Ridge Regression may serve as practical alternatives.

3.6. Limitations and Future Research Directions

Although the models performed exceptionally well on the Jena Climate dataset, several limitations should be considered. First, the dataset used in this study is location-specific, meaning the findings may not generalize to different climate zones. Future research could explore model performance on diverse climate datasets collected from various geographical regions. Additionally, while machine learning models demonstrated high accuracy, they lack the interpretability of physics-based meteorological models. Future work could investigate hybrid approaches that integrate machine learning with physical climate modeling. Another potential extension of this study involves deep learning architectures, such as Long Short-Term Memory (LSTM) networks or Transformer-based models. These architectures are capable of capturing temporal dependencies, making them well-suited for time-series forecasting. Furthermore, incorporating additional meteorological variables, such as precipitation and cloud cover, may enhance predictive performance. Lastly, an analysis of feature importance using techniques such as SHAP (SHapley Additive exPlanations) could provide deeper insights into the relationships between meteorological variables and temperature.

4. CONCLUSION

The results of this study demonstrate that machine learning models can achieve highly accurate temperature predictions when appropriately tuned and optimized. Through rigorous experimentation and hyperparameter tuning, it was observed that ensemble-based models, particularly Random Forest, Gradient Boosting, and XGBoost, exhibited the highest predictive performance with R^2 values approaching 1.0, confirming their ability to effectively model complex dependencies in meteorological data. The optimal hyperparameters for these models involved using 100 estimators and, in the case of Gradient Boosting and XGBoost, a learning rate of 0.1, striking a balance between learning efficiency and overfitting prevention. These findings indicate that boosting and bagging methods are highly effective for temperature prediction when a sufficiently large dataset is available.nThe analysis also highlighted the importance of feature scaling for certain models, particularly Support Vector Regression (SVR). The impact of different normalization techniques on SVR revealed that Min-Max scaling produced the highest R^2 score, while Robust Scaling led to a significant performance decline. This confirms that distance-based models such as SVR are highly sensitive to feature distribution, emphasizing the necessity of selecting appropriate scaling techniques in preprocessing pipelines. Furthermore, linear regression models such as Ridge and Lasso performed well with mild regularization, suggesting that the dataset retains a strong linear structure where simple models can still achieve high accuracy. Future work could focus on exploring deep learning architectures, integrating additional meteorological features, and testing the models across diverse climatic regions to ensure generalizability. These results provide a solid foundation for improving machine learning approaches in weather prediction and climate modeling, offering valuable insights for applications in meteorology, energy management, and environmental monitoring.

REFERENCES

[1] A.-L. Balogun *et al.*, “Assessing the potentials of digitalization as a tool for climate change adaptation and sustainable development in urban centres,” *Sustain. Cities Soc.*, vol. 53, p. 101888, 2020.
[2] S. M. Khan *et al.*, “A systematic review of disaster management systems: approaches, challenges, and future directions,” *Land*, vol. 12, no. 8, p. 1514, 2023.
[3] R. Meenal *et al.*, “Weather forecasting for renewable energy system: a review,” *Arch. Comput. Methods Eng.*, vol. 29, no. 5, pp. 2875–2891, 2022.
[4] S. K. Balasundram, R. R. Shamshiri, S. Sridhara, and N. Rizan, “The role of digital agriculture in mitigating climate change and ensuring food security: an overview,” *Sustainability*, vol. 15, no. 6, p. 5325, 2023.
[5] E. Stakhiv and B. Stewart, “Needs for climate information in support of decision-making in the water sector,” *Procedia Environ. Sci.*, vol. 1, pp. 102–119, 2010.
[6] I. E. Agbehadji, S. Schütte, M. Masinde, J. Botai, and T. Mabhaudhi, “Climate risks resilience development: A bibliometric analysis of climate-related early warning Systems in Southern Africa,” *Climate*, vol. 12, no. 1, p. 3, 2023.
[7] P. Sinha, M. Modani, S. Islam, M. Khare, and R. K. Srivastava, “Evolution of Weather and Climate Prediction Systems,” in *Mitigation and Adaptation Strategies Against Climate Change in Natural Systems*, Springer, pp. 243–265, 2025
[8] X. Bai, L. Zhang, Y. Feng, H. Yan, and Q. Mi, “Multivariate temperature prediction model based on CNN-BiLSTM and RandomForest,” *J. Supercomput.*, vol. 81, no. 1, p. 162, 2025.
[9] A. G. Di Stefano, M. Ruta, and G. Masera, “Advanced digital tools for data-informed and performance-driven design: a review of building energy consumption forecasting models based on machine learning,” *Appl. Sci.*, vol. 13, no. 24, p. 12981, 2023.
[10] D. R. Liyanage, K. Hewage, S. A. Hussain, F. Razi, and R. Sadiq, “Climate adaptation of existing buildings: A critical review

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2672−2680
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.7114

on planning energy retrofit strategies for future climate," *Renew. Sustain. Energy Rev.*, vol. 199, p. 114476, 2024.
[11] M. Hosseini, K. Javanroodi, and V. M. Nik, "High-resolution impact assessment of climate change on building energy performance considering extreme weather events and microclimate--Investigating variations in indoor thermal comfort and degree-days," *Sustain. Cities Soc.*, vol. 78, p. 103634, 2022.
[12] D. Markovics and M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," *Renew. Sustain. Energy Rev.*, vol. 161, p. 112364, 2022.
[13] S. He, X. Li, T. DelSole, P. Ravikumar, and A. Banerjee, "Sub-seasonal climate forecasting via machine learning: Challenges, analysis, and advances," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 1, pp. 169–177, 2021
[14] H. Shaiba *et al.*, "Weather Forecasting Prediction Using Ensemble Machine Learning for Big Data Applications.," *Comput. Mater. \& Contin.*, vol. 73, no. 2, 2022.
[15] A. Mahabub, A.-Z. S. Bin Habib, M. R. H. Mondal, S. Bharati, and P. Podder, "Effectiveness of ensemble machine learning algorithms in weather forecasting of Bangladesh," in *Innovations in Bio-Inspired Computing and Applications: Proceedings of the 11th International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA 2020) held during December 16-18, 2020 11*, pp. 267–277, 2021
[16] A. A. H. Lateko, H.-T. Yang, and C.-M. Huang, "Short-term PV power forecasting using a regression-based ensemble method," *Energies*, vol. 15, no. 11, p. 4171, 2022.
[17] A. Sekertekin, M. Bilgili, N. Arslan, A. Yildirim, K. Celebi, and A. Ozbek, "Short-term air temperature prediction by adaptive neuro-fuzzy inference system (ANFIS) and long short-term memory (LSTM) network," *Meteorol. Atmos. Phys.*, vol. 133, pp. 943–959, 2021.
[18] M. Yu, F. Xu, W. Hu, J. Sun, and G. Cervone, "Using Long Short-Term Memory (LSTM) and Internet of Things (IoT) for localized surface temperature forecasting in an urban environment," *IEEE Access*, vol. 9, pp. 137406–137418, 2021.
[19] E. Haque, S. Tabassum, and E. Hossain, "A comparative analysis of deep neural networks for hourly temperature forecasting," *IEEE Access*, vol. 9, pp. 160646–160660, 2021.
[20] M. Jansen, "BGC Jena Weather Station Dataset (2017-2024)." Kaggle, 2024.