Machine Learning Models with Statistical Analysis Techniques for Forecasting Wind Turbines Scada Systems Measurement

Mona Ahmed Yassen[1,2,*], El-Sayed M. El-kenawy[3,4,5], Mohamed Gamal Abdel-Fattah[2], Islam Ismael[2], Hossam El.Deen Salah Mostafa[2]

[1]Faculty of Artificial Intelligence, Hours University, Egypt

[2]Department of Electronics and Communications Engineering, Faculty of Engineering, Mansoura, 35516, Egypt

[3]School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT), Bahrain Polytechnic, PO Box 33349, Isa Town, Bahrain

[4]Applied Science Research Center. Applied Science Private University, Amman, Jordan

[5]Jadara University Research Center, Jadara University, Jordan

Emails: Monagaffer@std.mans.edu.eg; kenawy@ieee.org; eng.mo.gamal@mans.edu.eg; s islam_m@mans.edu.eg; hossam_moustafa@mans.edu.eg

Abstract

Wind energy is one of the fastest-growing sustainable, clean, and renewable sources, attracting significant attention and investment from many countries. However, given the substantial capital investment required for wind power plants, understanding the proposed plants' performance becomes critical before implementation. This assessment is most effectively conducted using refined wind power predictability models and precise wind velocity data. Accurate wind forecasts are essential for informed decision-making and effective wind energy utilization. In this study, three advanced Machine Learning (ML) regression methods were applied to the TNWind dataset to predict the power output of wind turbines. The dataset variables included date and time (measured at 10-minute intervals), low-voltage active power (in kW), wind speed (in m/s), the theoretical wind power curve (in kWh), and wind direction. To predict wind power output, six supervised ML models were trained, including Random Forest Regressor (RF), Extreme Gradient Boosting Regressor (XGB), Gradient Boosting Regressor (GB), Support Vector Machine Regressor (SVR), K-Neighbors Regressor (KN), and Linear Regressor. The analysis revealed that the Random Forest model outperformed the others, achieving exceptional performance metrics: an R^2 value of 0.97, an MAE of 0.17 and an MSE of 0.07. The analysis to identify the outcomes for wind power generation from machine learning proves that renewable energies are more capable and are a lucrative investment.

Keywords: Renewable Energy; Machine Learning; Wind; XGB; RF; Scada

1 Introduction

To effectively address the challenges arising from the remarkable growth in Renewable Energy (RE) production [1] [2], particularly the significant fluctuations in energy output, it is essential to enhance the precision of forecasting methods. Forecasting also plays a vital role in several fields of activity. Optimizing user activity scheduling is achieved by estimating the additional energy requirement of paramount importance in managing conspicuous consumption. One of the most important features is the capacity to forecast when consumers demand more or less energy so that the consumer can thus conserve energy use and cost. This preventive strategy is favorable to the individual consumer while at the same time supporting the balancing of the energy framework. With it, one can deploy the reserve power plants for usage during low renewable generation or even

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

during events of high demand. Accurate forecasting aids in developing optimized battery loading schemes [3] [4], ensuring that energy storage systems are charged and discharged appropriately to balance supply and demand. Without reliable forecasting, procuring wind energy and other variable renewable energies within the current grid system would be far more challenging. Since energy sources are very unpredictable, especially given the subjectivity of some sources, the grid's stability and efficiency are essential. Out of the many possible approaches to wind power integration, the fluctuating nature of wind requires strong forecasting techniques. Refining these forecasts for incorporating wind energy is the only positive and effective means to ensure the stability of the future energy industry. Thus, the question arises of managing problems connected with high fluctuations in energy rates about record RE generation growth. Optimal UFOE allows for the required co-ordination of the user timetable with power generation, thus improving the performance of the emergency power station for any fluctuations in demand. The size of the output power and the potential and realistic storage of the excess energy produced during optimum wind conditions must be accurately evaluated. These enhancements enhance the implementation of sound battery charging practices, enabling the integration of wind power plants into the grid. Last but not least, the reliability of the specified forecasting systems defines the possibilities of working through all the difficulties connected with using wind energy and developing practical, efficient options in the power industry. Mitigating these challenges guarantees a sufficient and reliable power supply that supports the theme of sustainable energy. Energy consumption has historically escalated alongside economic growth. As nations industrialize and GDP per capita rises, global energy consumption follows suit. Projections indicate that annual energy consumption will increase by approximately 0.8% by 2030, though this growth rate is expected to slow to around 0.1% by 2050 [5] [6]. Over the past decade, the focus has shifted towards RE utilization, resulting in significant changes in the consumption of conventional energy sources, especially hydrocarbons. Fossil fuels such as coal, oil, and natural gas have historically been the backbone of energy provision worldwide [7]. However, as renewable energy resources mature and become more affordable, they increasingly replace conventional energy sources. Wind energy [8], in particular, is emerging as a promising low-carbon alternative within modern energy grid systems. Similarly, a solar system that produces electricity without producing poisonous greenhouse gases is a good solution to the problem. However, the fluctuations in the flow of winds make undertaking and implementing large-scale wind technologies challenging. The need for renewable power sources is growing daily, so building better and more reliable models is no longer optional. Specifically, it is critical to predict wind speed and the amount of electricity produced by wind resources [9]. Such forecasts ensure that energy supply and demand are well coordinated and help determine where the power needs to go in the grid system. Artificial Intelligence (AI) [10], which can be considered a branch of computer science and engineering, aims to build intelligent systems performing tasks similar to those people execute. AI is a tool that tries to model mental states and procedures to create efficient tactics for various problems in science and real life. Its practical uses are becoming more apparent across businesses like RE. Several groundbreaking approaches are emerging from AI specifically, from Machine Learning (ML) to address energy problems and improve the effectiveness and eco-friendliness of power systems. These can go a long way to revolutionizing energy management and resource optimization while the world's energy needs keep rising. Such technology makes it possible for electronics, which the human brain resembles, to emulate the human brain, thus resulting in efficiency in energy source control. The division of AI models into a black box, white box, and grey box solely enhance the accuracy and performance of the energy control systems. Wind forecasting: The process usually involves estimating the anticipated energy generation of a particular wind power station in kilowatts or megawatts, including the total gross capacity of the wind farm. Current methods combine statistical and physical approaches, with the reliability of wind forecasts often tailored to specific areas. Wind forecasting is broadly categorized into four types: numerical methods, statistical approaches, physical modeling, and hybrid techniques [11] [12]. Numerical methods involving elaborate computer models are used to study changes in wind data in time and space. Physical forecasting was based on the Numerical Weather Prediction (NWP) models, which focused on effective wind speed and considered meteorological factors influencing wind farms. It does so in a way that directly converts predicted wind speeds into possible energy outputs, recasting estimates of power generation in a more precise light [13][14] On the other hand, statistical models work based on historical wind measurements and perform stochastic computations, like decision trees and multilayer perceptron, to estimate future wind energy generation. These methods use empirical data to give correct prediction results. Together, these approaches offer a range of flexible solutions for modeling wind energy, each with advantages and applications within the RE sector. In this paper, we propose an ML model for wind power prediction. This study addresses thirteen critical questions related to wind energy conversion systems. Based on the identified gaps, thirteen important questions concerning wind energy conversion systems are posed in this study. When used together, these techniques increase the reliability of the models used for forecasts of wind's future capacity for energy production, which is crucial for integrating RE resources into today's utility grids. The structure of this paper is as follows: Section 2 discusses

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

works similar to the work being conducted for the present paper, and section 3 outlines materials used in the research. Section 4 provides a comprehensive evaluation and discussion of the experimental results. Finally, conclusions from the work are presented in Section 5.

2 Related Works

Modern developments in ML and AI have progressively been applied to improve RE solutions, especially wind power prognostic models [15][16][17]. Many authors have employed numerous methods to assess and forecast wind power output, often using kNN algorithms with significant inputs, including meteorological parameters. Bonfil et al. [18] created a novel model based on SVR for wind speed estimation in wind farms, emphasizing its importance for maximizing wind energy generation. Similarly, Mi et al. [19] proposed a method combining wavelet transforms with an extreme learning machine and an outlier correction algorithm, significantly increasing wind speed forecast accuracy. In related research, Sareen et al. [20] demonstrated the feasibility of deep learning techniques by imputing wind speeds in five Indian cities using a bidirectional LSTM network. This approach revealed the intricate structure of wind data signals and showed better predictive accuracy than conventional methods. Abdoos et al. [21] applied the variational mode decomposition method for deterministic wind power forecasts. Ye et al. [22] developed an ultra-short-term wind power prediction model using the random forest algorithm coupled with trending feature similarity and historical data matching. Zhao et al. [23] utilized LightGBM to predict wind power, transforming wind power data into components using NeuralProphet. In contrast, Liao et al. [24] used the mutual information coefficient to screen input features. Liu et al. [25] proposed an empirical mode decomposition-LSTM-AARIMA-based model to enhance wind speed forecast accuracy. Their method improved performance by focusing on high-frequency residuals and subsequences. For short-term wind speed prediction, studies [26] [27] [28] explored wavelet-based SVR, random forests, CNNs, and twin SVR. These studies also developed a twin SVM and adaptive thresholding method to identify gearbox anomalies in wind turbines. Using genetic programming, Zameer et al. [29] introduced a short-term wind power forecasting technique based on multiset L-L externally matched neural networks. Li et al. [30] presented an SVM-based approach for wind power prediction, employing data mining techniques to analyze the correlation between wind speed and power output while correcting erroneous initial readings. Mahmoud et al. [31] predicted wind power using a hybrid intelligent system combining self-adaptive evolutionary extreme learning and extreme learning machines. Their approach achieved top results using Extra Trees, LightGBM, and Swiss wind speed data ensemble methods. For SCADA systems, the best RMSE was achieved with a light gradient boosting machine and AdaBoost ensemble approaches. Malakouti and Seyed Matin [32] found that gradient-boosting regression provided better forecasts than benchmark models. At the same time, random forests outperformed AdaBoost and K Neighbors in estimating wind speed and power production at a Texas wind farm. Rashid et al. [33] used French SCADA data with the RF Regressor algorithm, achieving rapid learning, approximate results, and minimal overfitting. For load forecasting, Guot et al. [34] evaluated three ML techniques LSTM, RF Regression, and SVR. While RFR proved simple to implement, fast to train, and marginally resistant to overfitting, it significantly outperformed SVR for short-term load forecasting. Capacity factor, a critical measure of wind farm performance, reflects the facility's expected and actual output. This work employed MAE and MSE in assessing model reliability; site-specific attributes and input imbalance caused oscillation in the two parameters. The datasets used in these studies are region-specific, which makes a clear case for coupled dependence between inputs and outputs in power output estimates. This research focuses on the analysis of wind turbine performance prediction with a special focus on Turkey's environment. Last but not least, a tabular comparison of the models used for wind power forecasting is presented in Table 1.

3 Methods

This section explains the datasets, data transformation method, architecture of the models, and the assessment metrics used in the wind forecasting investigation. Nonetheless, the readers of this paper can find out the error rates and the statistical analysis of the efficiency of the models. Based on the results obtained in this study, there is every indication that the models in use are highly accurate and that dependable forecasts could be provided due to the application of diverse techniques. Nevertheless, the preprocessing step is the initial phase of the data analysis procedure necessary for training machine learning models. As shown in the block diagram

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

Table 1: An overview of the ML and AI methods for predicting wind speed and power.

Authors	Approaches/Models Employed	The primary objective
Bonfil et al. [18]	Support Vector Regression (SVR)	Wind speed prediction for wind farms
Mi et al. [19]	Wavelet, Extreme ML, Outlier Correction Algorithm	Wind speed prediction
Sareen et al. [20]	Bidirectional Long Short-Term Memory (LSTM)	Wind speed prediction for five Indian cities
Abdoos et al. [21]	Variational Mode Decomposition	Deterministic wind power forecast
Ye et al. [22]	Random Forest, Historical Data Matching	Ultra-short-term wind power prediction
Zhao et al. [23]	LightGBM, NeuralProphet Decomposition	Robust wind power prediction
Liao et al. [24]	Mutual Information Coefficient	Filtering input features for wind power prediction
Li et al. [30]	Support Vector Machine (SVM)	Wind power prediction using data mining
Malakouti et al. [32]	ML Algorithms (Extra Tree, Gradient Boosting, Ada Boost)	SCADA system power production prediction
Singh & Rizwan [35]	Gradient Boosting Regression	Estimating wind power generation
Malakouti & Seyed Matin [36]	Random Forest (RF), Ada Boost, K Neighbors	Wind speed and power prediction for Texas wind farms
Rashid et al. [33]	RF Regressor	Wind turbine performance forecasting using French SCADA data

in Figure 1, this complex process comprises the following critical stages. Special attention is paid to the choice of target variables and key characteristics for which inclusion into the model is critical. Data normalization is the next step. It scales the numerical values while maintaining the differences in the range. It makes the training process more effective and helps produce a better model. Also, essential steps for preparing data to apply a particular machine learning include preprocessing data to fit into the format required by the specified machine learning. The features included in the dataset are as follows:

- **Date/Time (10-minute intervals):** This feature measures duration in terms of 10 minutes past the hour. It makes it possible to conduct frequent and more comprehensive assessments of deviations within short-term activities.
- **LV Active Power (kW):** Expressed in kilowatts (kW), this figure characterizes the amount of electrical power output of the turbine at a given point in time. It gives relevant data on the efficiency of the turbine and the prevailing wind climate at the locality.
- **Wind Speed (m/s):** This parameter reflects wind speed at turbine height and is measured in meters per second (m/s). The wind speed is a significant predictor of power generation, and the wind speed level determines the effectiveness of wind energy. The maximum design-rated power of the turbine, the speed of the wind at which the turbine operates, and the power curve data in kilowatt hours supplied by the manufacturer are crucial in this respect.
- **Wind Direction (°):** This parameter defines wind direction at turbine height and must be an integer value between 0 and 360 inclusive. Wind turbines are usually orientated with an optimum wind direction in mind, and wind capturing is highly dependent on this orientation.

Together, these features form a robust dataset that enables the training and testing of highly complex machine-learning algorithms essential for reliable wind forecasting.

Graph (a) in Figure 2 provides a comprehensive overview of the frequency distribution of wind speeds in the dataset. Evidently, from the data, the highest frequency is 7m/s, with most speeds ranging between 5 to 10m/s. First, I would like to draw attention to the fact that long tails are pointed upwards, which means that wind speed values are less frequent but more significant than the most frequent value of wind speed, contributing to a skewed distribution.

DOI: https://doi.org/10.54216/FPA.190205
Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

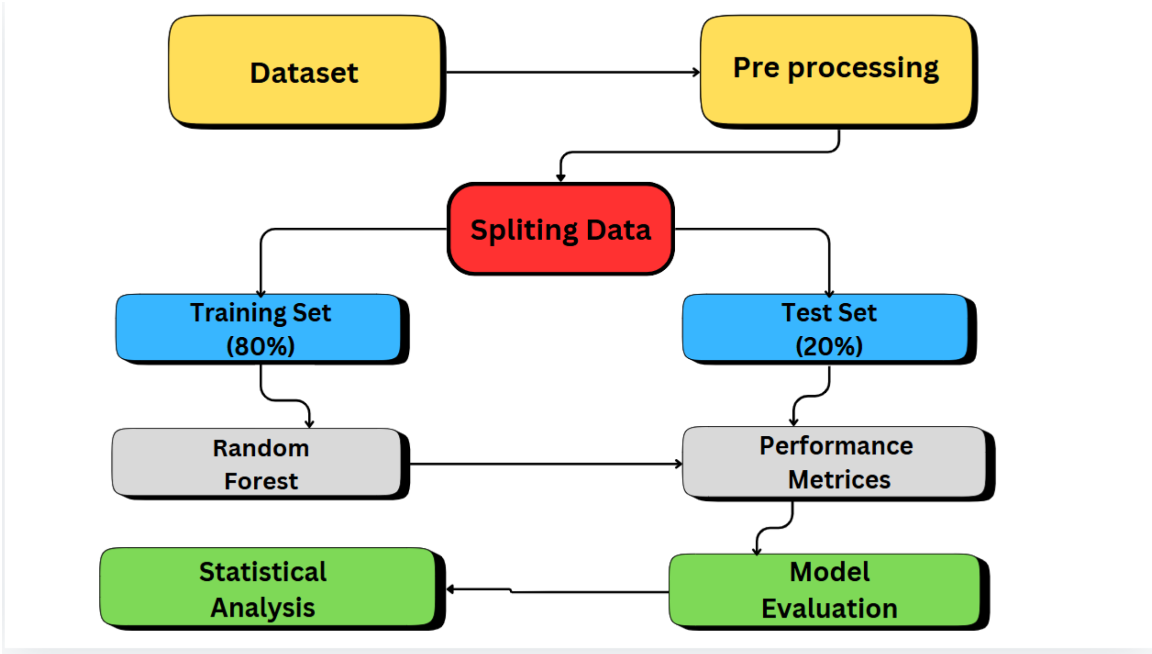
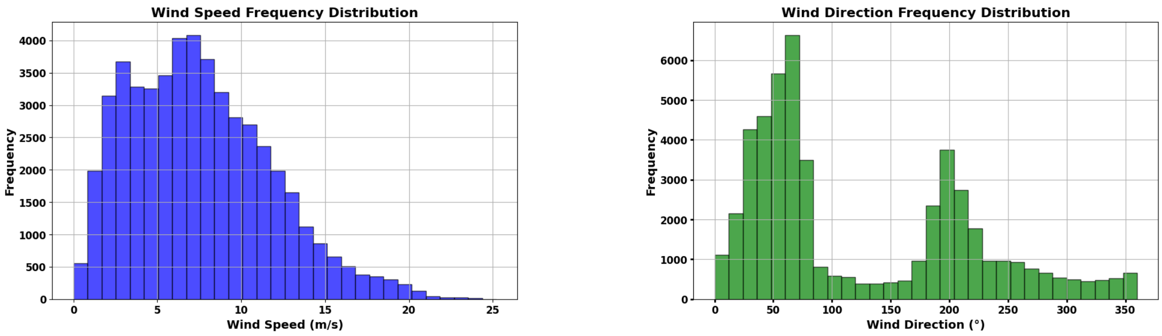


Figure 1: Block schematic of the suggested methodology.



(a) (b)

Figure 2: Distribution of Data

Moving to Graph (b), we see a detailed representation of wind directions, which are predominantly clustered in two key ranges: from 0° to 100° and from 180° to 230°. The most prevalent wind direction is approximately 60°, which boasts over 6,000 recorded occurrences, indicating a strong preference for this direction in the dataset. In contrast, winds blowing between 100° and 150° are relatively uncommon. A significant frequency is observed in the 180° to 230° range, indicating a robust pattern of wind activity in this sector. Meanwhile, wind directions from 250° to 350° appear much less frequent, suggesting limited wind activity. Figure 3 includes data encompassing all possible wind directions for testing and analysis, providing a holistic view of the wind patterns.

3.1 Developmental Mechanism

The last train-test split method tests models by feeding 80% of the data to the training set and 20% to the test set. This step also provides the opportunity to check the model's accuracy on new data, thus increasing its ability to generalize data points. Python, Google Colab, and sci-kit-learn are some of the tools used in data science that are so helpful. Python has lots of data analysis options. Google Colab is a comfortable online environment connected with Google Drive, accessible to GPUs and sci-kit-learn, rich in algorithms for classification and regression, clustering, etc. TensorFlow is used for deep learning, NumPy for numerical work,

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

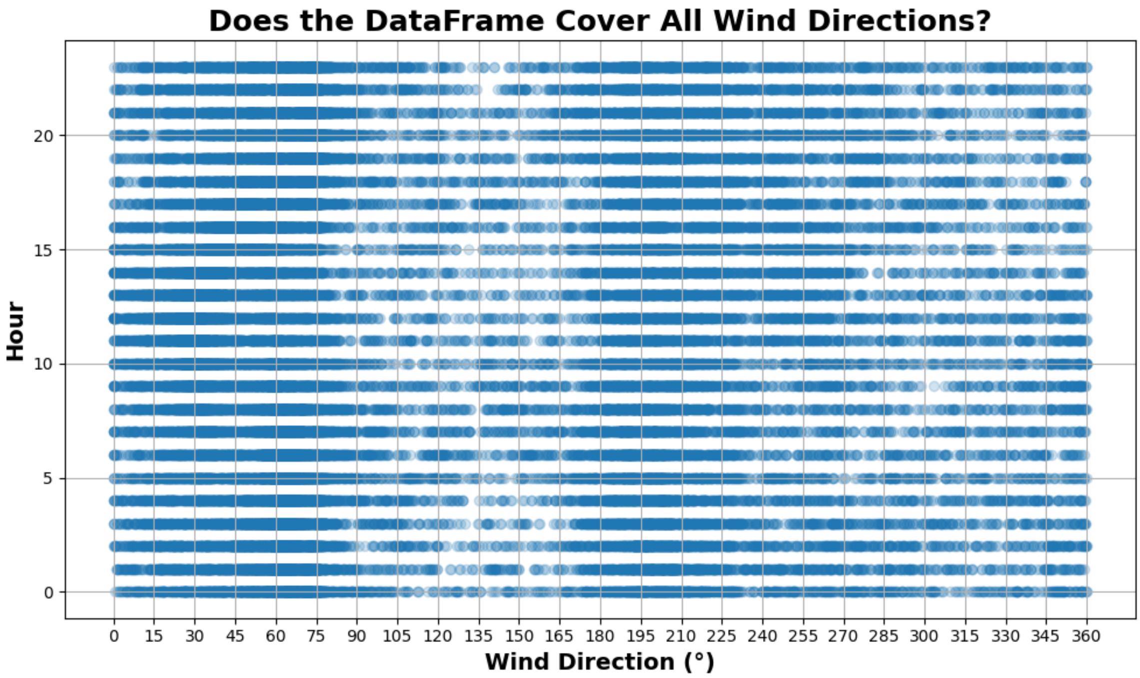


Figure 3: Data frame covers all wind direction

and Pandas for simple data computation and manipulation using data frames. All graphs are created using the Matplotlib library, and the Seaborn library is ideal for statistical graphs. Pandas are exceptional at handling and manipulating data frames and data analysis, while Matplotlib is quite helpful in plotting information and Seaborn for visual data analysis. All these tools work in a way that allows the intricate examination of a dataset and the building of the best machine-learning models.

4 Results and discussion

The results presented in Table 2 show the performance of six ML models: Random Forest Regressor (RF),

Table 2: Model performance metrics comparison

Model	RMSE	MAE	MBE	r	R2	RRMSE	NSE	WI
RF	0.179	0.069	0.002	0.983	0.967	1711.02	0.967	0.99
XGB	0.17	0.07	0.002	0.98	0.97	1640.33	0.97	0.99
GB	0.25	0.130	0.001	0.96	0.93	2411.71	0.93	0.98
SVR	0.250	0.100	0.022	0.96	0.93	2381.42	0.935	0.984
KN	0.177	0.066	0.002	0.984	0.968	1687.52	0.96	0.992
Linear Regression	0.306	0.168	0.0019	0.952	0.906	2912.35	0.906	0.975

Extreme Gradient Boosting Regressor (XGB), Gradient Boosting Regressor (GB), Support Vector Regressor (SVR), KNeighbors Regressor (KN), and Linear Regression. These models were evaluated using various metrics, including RMSE, MAE, MBE, correlation coefficient (r), R^2, RRMSE, NSE, and WI, formally defined in the provided equations as shown in Table 3.

Among these models, RF stands out as a robust method for numerical prediction. It constructs decision trees based on a random vector whose output values depend on this randomization. RF enhances stability by

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

Table 3: Statistical Performance Metrics and Their Equations

Metric	Equation
Root Mean Square Error (RMSE)	$\text{RMSE} = \sqrt{\frac{1}{s}\sum_{m=1}^{s}(x_m - \hat{x}_m)^2}$
Mean Absolute Error (MAE)	$\text{MAE} = \frac{1}{s}\sum_{m=1}^{s}\lvert x_m - \hat{x}_m\rvert$
Mean Bias Error (MBE)	$\text{MBE} = \frac{1}{s}\sum_{m=1}^{s}(\hat{x}_m - x_m)$
Correlation Coefficient (r)	$r = \frac{\sum_{m=1}^{s}(x_m - \overline{x})(\hat{x}_m - \overline{\hat{x}})}{\sqrt{\sum_{m=1}^{s}(x_m - \overline{x})^2}\sqrt{\sum_{m=1}^{s}(\hat{x}_m - \overline{\hat{x}})^2}}$
Coefficient of Determination (R²)	$R^2 = 1 - \frac{\sum_{m=1}^{s}(x_m - \hat{x}_m)^2}{\sum_{m=1}^{s}(x_m - \overline{x})^2}$
Relative Root Mean Square Error (RRMSE)	$\text{RRMSE} = \frac{\sqrt{\frac{1}{s}\sum_{m=1}^{s}(x_m - \hat{x}_m)^2}}{\overline{x}}$
Nash-Sutcliffe Efficiency (NSE)	$\text{NSE} = 1 - \frac{\sum_{m=1}^{s}(x_m - \hat{x}_m)^2}{\sum_{m=1}^{s}(x_m - \overline{x})^2}$
Willmott Index of Agreement (WI)	$\text{WI} = 1 - \frac{\sum_{m=1}^{s}(x_m - \hat{x}_m)^2}{\sum_{m=1}^{s}\left(\lvert\hat{x}_m - \overline{x}\rvert + \lvert x_m - \overline{x}\rvert\right)^2}$

averaging multiple decision trees, which reduces overfitting and lessens the significant variation associated with individual trees. Similarly, XGB is a practical, adaptable, and portable tool for time series prediction. It improves accuracy and robustness by employing decision trees and lagged features through ensemble learning. XGB's ability to handle missing data, model complex relationships, and capture temporal patterns makes it a precious model for time series data. The model's predictions are calculated by accumulating past predictions, as shown in Equation 1:

$$\hat{X}_i = \varphi(y_i) = \sum_{n=1}^{K} v_n(X_i),\ v_n \in V \tag{1}$$

SVR, a specialized field of Support Vector Machines (SVM) [37], uses kernel functions to address nonlinear problems by minimizing the total deviation between sample points and the hyperplane. On the other hand, KN is a simple yet effective technique that predicts outcomes based on the average of the target variable's k-nearest neighbors. However, it is sensitive to outliers and noise, relying on local patterns in the feature space [38]. Finally, Linear Regression [39] remains one of the simplest models, aiming to find the best-fitting line in the dataset. It minimizes the distance between each data point and the regression line, creating a linear relationship between the labels and features across the dataset. Table 4 illustrates the time each model took to run is included. XGB has the lowest average RMSE of 0.17, r=0.98, and the highest R2 =0.97. This

Table 4: Time taken by each model

Model	Time (seconds)
RF	48.6
XGB	1.6
GB	11.9
SVR	17.9
KN	0.332
Linear Regression	0.007

demonstrates its capacity to ensure that predictions are as close as possible to the actual values with minimal approximation; hence, it is a good candidate for predictive problems. On the other hand, the RF model is also relatively high performing, with values similar to XGB; thus, RMSE = 0.17 and R2 = 0.967. Although it is not as accurate as XGB, it is still quite viable for prediction tasks. As with the KN model, the present

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

study yielded a good measure of fit in terms of performance and accuracy with a low RMSE of 0.177 and correlation coefficient of r=0.984. However, it is slightly less effective than XGB and RF. Meanwhile, despite performing reasonably well, both GB and SVR models show higher errors (RMSE above 0.25), indicating they may need to generalize better in this context. Conversely, linear regression presents the worst results, with the maximum RMSE= 0.306 and the minimum R2 = 0.906, respectively, making this model less efficient thanthan other models. On the other hand, the execution times of models differ considerably. Among all the models, linear regression processes the data most quickly as it took only 0.0076 seconds; however, it needs to be more accurate for sophisticated tasks within which precision is critical. It also has a speed of 0.33 s, which is suitable for a search that needs quick results, considering that KN is suitable for such a search. XGB makes the model the most accurate with an accuracy of 96.15% and the second most efficient by taking a time of 1.67 secs only. In contrast, the RF model takes significantly longer (48.68 seconds), making it less suitable when speed is a priority. Both GB and SVR models take moderate times (11.98 and 17.95 seconds, respectively) but provide lower accuracy than RF and XGB, reducing their overall value. Comparing relative error and efficiency to the observed values, the RRMSE is the largest for linear regression (2912.35) due to the low accuracy of modeling. On the other hand, the lowest RRMSE values, about 1640 of XGB and RF, show better predictive outcomes concerning the data variation. Additionally, Nash-Sutcliffe Efficiency (NSE) values highlight the high effectiveness of XGB (0.970) and RF (0.967), as these values are close to 1, indicating excellent predictive power. Willmott's Index (WI) confirms the same trend, with values close to 1 for XGB (0.992), RF (0.991), and KN(0.992), showing strong agreement between predicted and observed values. Finally, the model choice would depend on the specific use case in practical applications. If speed is critical, KN or XGB would be preferable, whereas RF could be used when higher accuracy justifies the extra computation time. Figure 4 shows the relationship between predicted and actual values. Each plot contains a dashed red line representing the ideal line where actual values are expected to match predicted values. The closer the points are to this line, the better the model's performance. Predictive model performance is measured by the degree of similarity between the obtained forecasts and a specific reference value, depicted by the red line. A higher degree of collinearity between the predicted and actual data points indicates a more effective model. If the results of the actual run are depicted in Figure 4, it becomes apparent that the XGBoost (XGB) model is the highest performer among all the contenders. From the results as presented in Figure 4, one is absolved to infer that the model returns high-performing metrics such as the low RMSE of 0.17 (1.7%), $r = 0.98$, and $R^2 = 0.97$ indicating not only the accuracy of the predictions but also the ability of the model to predict accurately. Such statistically significant findings prove the XGB model is an invaluable tool for accurate and reliable prediction. The Random Forest (RF) model also has almost equal accuracy metrics on all the predictive functions, like the XGB algorithm. Nonetheless, the RF model is less predictive than the XGB model with RMSE = 0.17 and $R^2 = 0.967$. However, the RF model shows high stability and robustness in its performance since it also took the same time in the fifth iteration. Another good solution is the K-Nearest Neighbors model, for which three evaluation indicators were calculated: RMSE = 0.177, and the coefficient of determination $r = 0.984$. However, we observe that using all the distances in KNN is computationally expensive; therefore, KNN's efficiency rate is lower than the XGB and RF models, making it less practicable. In some cases, less accurate estimates provided by KNN may be acceptable, but they will not be appropriate in more accurate applications. Although Gradient Boosting (GB) and Support Vector Regression (SVR) models have the least number of trees, they report higher RMSE values of 0.38 and 0.30, respectively, meaning a generalization problem exists. Likewise, using the Linear Regression model, the obtained RMSE is equal to 0.306, and $R^2 = 0.906$, confirming that the application of today's more advanced algorithms, such as XGB and RF, outperforms this simple Linear Regression model. While comparing the speed of the models, one must always consider how the time taken to process affects functionality. Linear Regression again takes very little time to compile and complete the analyses; 0.0076 seconds were used. However, it also means that it needs to be more accurate and apt for use in cases where expectations must be predicted accurately and in detail. The model we developed, the Decision Tree model, is more accurate than the former, with an accuracy of about 87%, while the KNN model also performs well within 0.33 sec. The difference is stunning in the XGB model with 90% accuracy and a speed of 1.67 sec. Its high accuracy and speed make it suitable for users who predict predictions in the short term. On the other hand, there is the RF model, which is slower than the other models, with the processing time being 48.68 seconds on average. It earns a slightly lower recall rate of 79.97% The GB and SVR models have reasonable run times with an execution time of 11.98 seconds and 17.95 seconds for GB and SVR, respectively. However, their accuracy could be better than that of XGB and RF; therefore, they are not so beneficial for requiring precise forecasts. From the above results, the XGB model is the most balanced, accurate, and time-efficient of the models under consideration. It is, therefore, suitable for applications requiring accurate and real-time prediction.

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

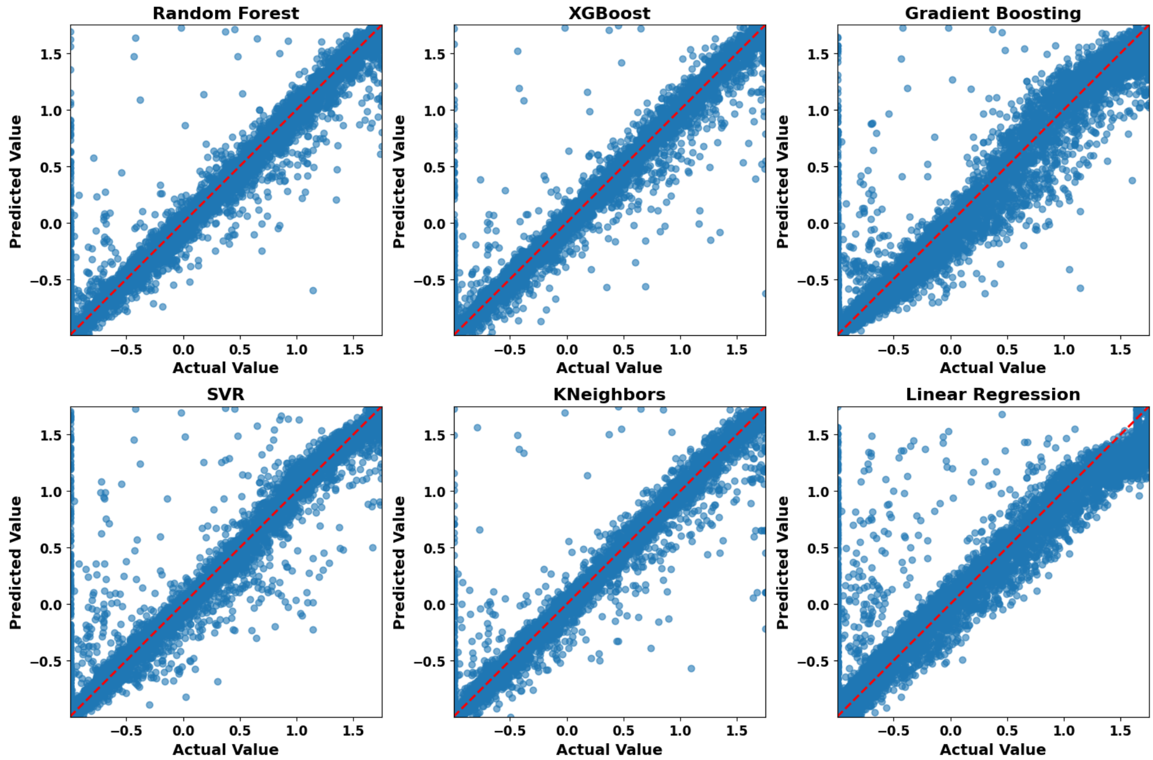


Figure 4: Relationship between Actual and Predicted Values.

Relative Root Mean Square Error (RRMSE) can also be used to measure the modeling efficiency of this dataset. Linear regression, however, seemed not fit to make estimates of the AHI because of the high RRMSE of 2912.35, which pointed to the ineffectiveness of this technique. Thus, there is great potential in the XGBoost (XGB) and Random Forest (RF) models, as their RRMSE values reach slightly above 1640. These results show that these models can better model these signals than linear regression. The K- Nearest Neighbors (KNN) model is among the models that take less time to process the data. The time it takes is 0.33sec. Such capabilities of a fast response make it applicable to applications that require quick responses. The proposed XGB model is slightly time efficient and out of the Ordinary in terms of accuracy; it completes in 1.67 seconds. While at times faster than its traditional counterpart in some scenarios, the RF model can be considerably slower at other times, which may be problematic, especially in a time-sensitive situation. Overall, the overall performance of the RF model shows the lowest RMSE of 0.1 and the highest R^2 of 0.967. This marks its ability to predict values accurately, though slightly less accurately than the XGB model. A good indication of the practical usability of the KNN model is demonstrated by RMSE = 0.177 and $r = 0.984$ observed for the current dataset. XGB and RF models provided greater accuracy and reliability than LASSO and EN models, though they will take longer to perform cross-validation. The GB and SVR models yield higher values of RMSE than 0.25, which suggests that the models generally overfit. Mean Absolute Error, abbreviated as MAAE, has been traditionally used when measuring the accuracy of the forecasts; Root Mean Square Error, abbreviated as RMSE, quantifies the mean square error of the observed and forecasted data. A higher RMSE value defines the more significant error variance and means that the model could underestimate or overestimate the training set. Linear regression has the highest RMSE value = 0.306 but has the second best $R^2 = 0,906$. Although it is essential and not complex, it takes a shorter time to analyze data, taking about 0.0076 secs. However, being unable to perform complex operations, they could be more helpful in performing complex tasks. In contrast, the KNN model considered in this study is computationally efficient, having taken an average of 0.33 seconds to compute. The efficiency of the KNN model is ideal for applications where quick decisions need to be made. The evaluated XGB model offers an almost ideal mean square error of 0.017 while only taking 1.67 seconds to execute its role, which also serves as a high confidence in its capacity. At the same time, the RF model needs about 48.68 sec to process the data, which is relatively slow and cannot be used in emergencies. GB and SVR models are computed within acceptable time, less than 12 seconds for GB and nearly 18 seconds for SVR, but they could be more useful due to lesser accuracy. Linear regression yields an RRMSE of 2912.35, which could be more efficient for this dataset and impractical. Hence, the XGB and RF models are even more appropriate for this dataset since they have superior accuracy and much lower error than the GB and SVR

DOI: https://doi.org/10.54216/FPA.190205
Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

models. However, for taste applications where immediate decisions need to be made, the KNN model's time taken to train is a plus. The NSE (Nash Sutcliffe Efficiency) further supports the high accuracy of selected models, of which XGB and RF were 0.970 and 0.967, respectively. That kind of robust performance increases confidence in the relevance of these models. In addition, performance measures estimated by the formula given in this study, along with the C Coefficient, also support these positive trends as the value of C falls in the range of ideal concordance, i.e., 1. Similarly, the Willmott Index (WI) Provides a backup for this, with values close to 1. In particular, therefore, the values acquired for the study models were WI of 0.992 for the XGB model, WI of 0.991 for the RF model, and WI of 0.992 for the KNN model. These high indices suggest that the model's predictions closely capture the sample figures obtained.

4.1 Satisical Result Analysis

The heatmap in Figure 5 uses a dynamic color gradient to heatmap the relationship between the different parameters in our dataset. Every color in this gradient is chosen deliberately because this component illustrates the degree of association of the parameters, with darker tones symbolizing stronger association. This diagram helps the analysts comprehend the connection and interactional properties of effects between various aspects.

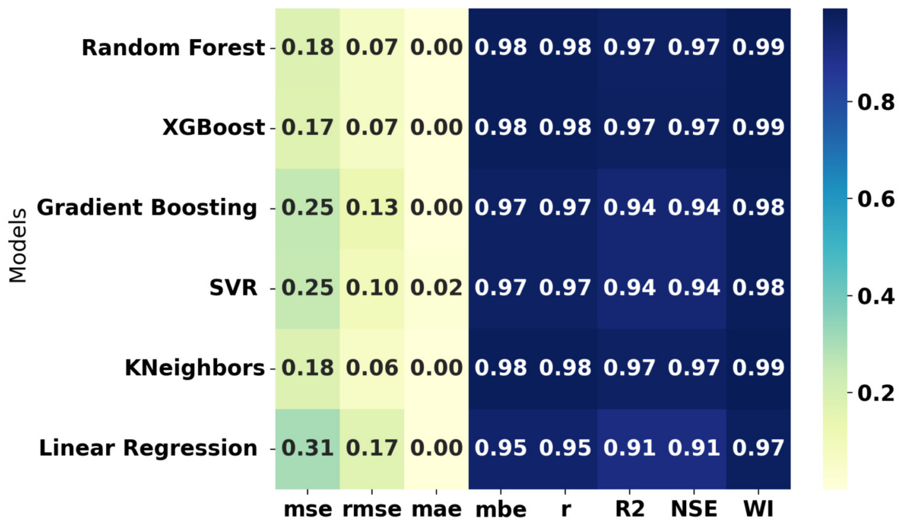


Figure 5: A heatmap showing how various ML models' efficiency is correlated with their features.

Employing this heatmap, we methodically examined variable interrelationships at the micro level of analysis. We identified multicollinearity Heatmaps—a behavior that occurs when two or more predictor features in models are highly correlated. That identification was eased by the heatmap, which made the condition analysis of the data more accessible. It is always important to be aware of Heatmap issues to allow for correction should this impact regression analysis or generate misleading information. That is why this study is significant. This paper recognized several components that significantly interact with each other, thus offering a better understanding of the dataset. In short, this enabled perspective seemed h, which helped appreciate details in data and orientations of parameters, which helped to provide better decisions or orientations for the following experiments. These are rich findings since greater scrutiny reveals additional nuances of the results and their interconnections. However, the radar chart presented in Figure 6 is an effective tool to present metrics to compare multiple forecasting algorithms. Analyzing the radar chart is less time-consuming and complicated than analyzing individual bar charts and simultaneously enables evaluations of each algorithm's strengths and weaknesses.

This chart compares several vital algorithms, including Linear Regression, Random Forest (RF), eXtreme Gradient Boosting (XGB), Gradient Boosting (GB), Support Vector Regression (SVR), and K-Nearest Neighbors (KNN). The chart visually emphasizes how each algorithm performs differently across various metrics. Graphic 6 provides a useful comparative overview for selecting the most suitable algorithm for specific forecasting goals. This approach simplifies determining which algorithms outperform others in accuracy or efficiency for the given analysis. Moreover, this analysis is complemented by accuracy verification forecasts of the ML models, presented in parallel in Figure 7, further enriching the assessment.

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

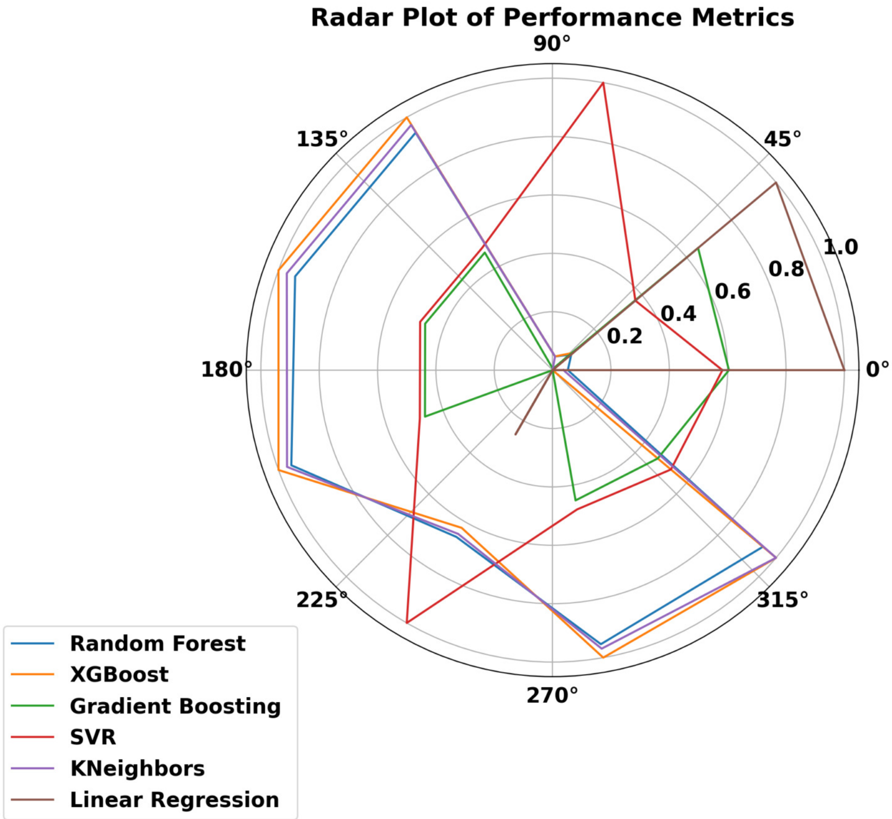


Figure 6: A radar chart comparing the performance of various models across multiple metrics.

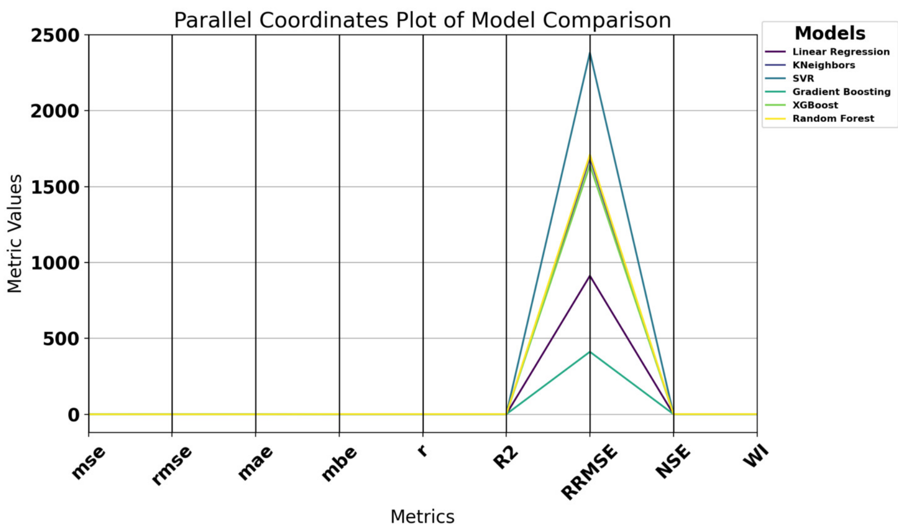


Figure 7: A chart presenting six metrics (e.g., MSE, RMSE, MAE) on the horizontal axis and their performance levels on the vertical axis.

The statistical significance of the models was meticulously evaluated using two distinct tests: the Wilcoxon signed-rank test and the Analysis of Variance (ANOVA) test. This study explicitly used the ANOVA test to assess the effects of various preprocessing methods on the given models. The comprehensive results from this test are summarized in Table 5, organized into three primary sections: Total, Effective, and Treatment.

In the treatment section of the ANOVA analysis, we observe five degrees of freedom (DF) alongside a mean square (MS) of 0.01793 and a sum of squares for treatment (SS) equal to 0.08964. The F statistic obtained, denoted as $F(5, 54) = 75.86$, is accompanied by a remarkably low P-value of less than 0.0001. These findings

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

Table 5: The prediction outcomes were evaluated using the analysis of variance (ANOVA) test.

ANOVA	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.08964	5	0.01793	F (5, 54) = 75.86	P ¡ 0.0001
Residual (within columns)	0.01276	54	0.0002363		
Total	0.1024	59			

indicate a significant statistical difference between the various preprocessing methods, suggesting that at least one of these techniques has a noteworthy influence on the model's performance. The residual section highlights the discrepancies between the expected and actual outcomes. This section represents the unexplained variance or error associated with each preprocessing method. The sum of squares for the residual (Residual SS) amounts to 0.01276, with 54 degrees of freedom and an MS value of 0.0002363. This part of the analysis assesses the variance not attributable to the preprocessing methods, thereby emphasizing potential variability within the data or limitations imposed by the model. Moreover, the total section encapsulates the overall variance in the predictive model's performance. This is achieved by combining the residual and treatment components, resulting in a Total Sum of Squares (TSS) of 0.1024, based on 59 degrees of freedom. This figure underscores the extent of variance in the data that can be explained by applying these preliminary processing methods, which is a significant finding of our analysis. Additionally, the Wilcoxon signed-rank test was conducted to explore the data further, and its findings are detailed in Table 6. According to the results from this test,

Table 6: The model selection outcomes were analyzed according to the Wilcoxon signed-rank test.

	RF	XGB	GB	SVR	KN	Linear Regression
Theoretical Median	0	0	0	0	0	0
Actual Median	0.069	0.07	0.13	0.1	0.06	0.168
Number of Values	10	10	10	10	10	10
Wilcoxon Signed Rank Test						
Sum of Signed Ranks (W)	55	55	55	55	55	55
Sum of Positive Ranks	55	55	55	55	55	55
Sum of Negative Ranks	0	0	0	0	0	0
P Value (two-tailed)	0.002	0.002	0.002	0.002	0.002	0.002
Exact or Estimate?	Exact	Exact	Exact	Exact	Exact	Exact
P Value Summary	**	**	**	**	**	**
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes	Yes	Yes
How Big is the Discrepancy?	0.069	0.07	0.13	0.1	0.06	0.168

the methodologies employed for predicting the stability of the intelligent grid reveal significant differences when evaluated against various feature selection procedures that match. The theoretical mean value of zero implies no anticipated difference when comparing the medians of the paired procedures involved in the test. However, the results demonstrate noticeable disparities between the medians of diverse models, including RF, XGB, GB, and linear regression. These differences highlight variability in direction and the magnitude of the various methods used, underscoring the importance of the research. The analysis rejects the null hypothesis, supported by a two-tailed P-value of 0.002, indicating significant differences among the compared methods. These findings are of significant importance in statistical analysis and model evaluation. Figure 8 features four distinct charts that thoroughly examine the effectiveness of ML models in predicting energy usage.

The first one that must be discussed is the residual error chart, which has valuable information in confirming or rejecting the models. The second is the homoscedasticity chart, which shows how much spread the residuals in the model have for differing amounts of energy usage. The third graph under discussion is the Quantile

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

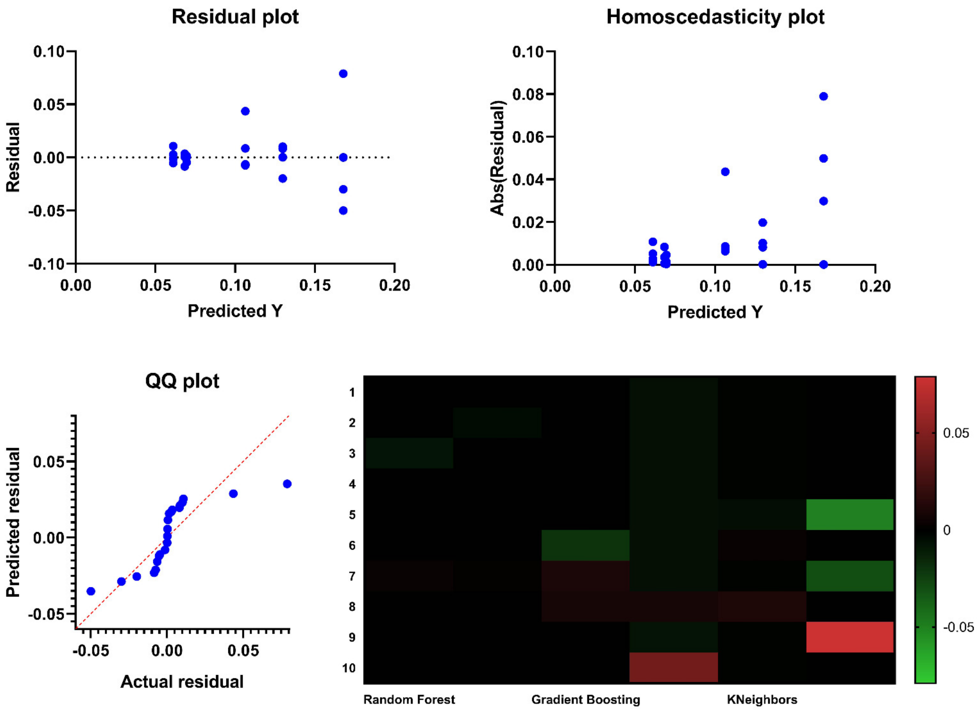


Figure 8: Presenting the prediction results of the proposed enhanced ML model.

plot, mainly used to compare the actual and modeled residuals. This process aids in the identification of sources of systematic error, which might be realized if the model was employed in calculations. Finally, the Heatmap compares the performance of all the developed models and demonstrates the stability of RF Heatmap algorithms. The following models present considerably lower associated error rates. These figures help compare how accurately these models can predict total energy consumption. As such, the results show that the models can enhance the predictability of energy use estimates.

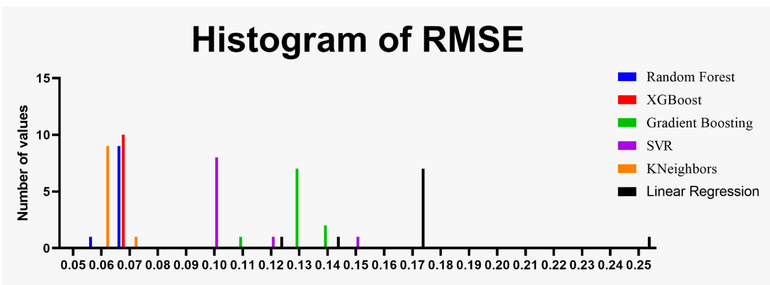


Figure 9: Histogram of RMSE values obtained using the suggested optimization technique.

In contrast to various alternative methods, the prediction errors associated with the proposed methodology are depicted in Figures 9 and 10. These figures illustrate how the suggested methodology effectively captures the forecasted wind energy generation pattern. When comparing the different models, stability and predictive performance variations are observed. These differences provide valuable insights into each model's comparative advantages and disadvantages. The obtained low coefficient of variation (CV) values, CV = 2.366% for XGB and CV = 4.512% for RF, indicate high consistency in the models' results. These findings suggest that the respective models produce minimal variations around the mean, making them reliable solutions for various predictive problems.

Interestingly, XGB also demonstrates very low variance, with relatively low standard deviations in test set errors, further enhancing its stability. This portrays the model as highly suitable for scenarios requiring predictability and consistency. However, linear regression produces the highest cumulative prediction total error of 1.679 and a CV of 19.46%. This increased variability indicates a lower level of certainty, particularly

DOI: https://doi.org/10.54216/FPA.190205
Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

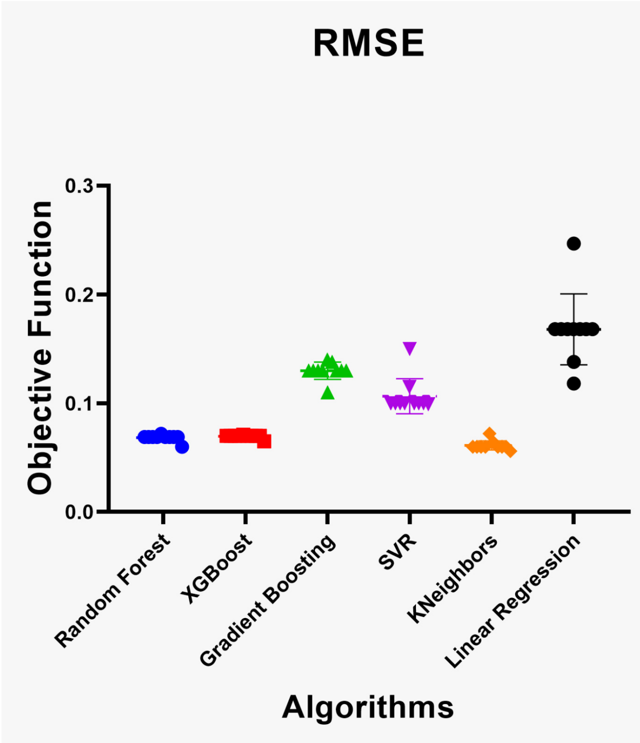


Figure 10: The results obtained by the suggested optimization technique's RMSE values.

in situations demanding precision, making linear regression less reliable for such applications. Delving deeper into the models' distributional characteristics, we find that the skewness and kurtosis values provide essential insights. The RF and XGB models demonstrate negative skewness, indicating that most predictions cluster towards the higher end of the value spectrum. On the other hand, models such as SVR and linear regression show a positive skew, suggesting that their predictions tend to concentrate at the lower end. Coupled with significant kurtosis levels, we find a propensity for producing extreme prediction outcomes. These statistical findings highlight the critical importance of carefully selecting and fine-tuning predictive models based on the inherent properties of the data. Such an approach is essential for enhancing the accuracy and stability of prediction tasks, as further illustrated in Table 7.

5 Conclusions

Since wind energy availability cannot be guaranteed, predicting how much wind energy will likely be available at a particular location is essential. The primary objective of this study was to propose models capable of accurately estimating the wind power generation of wind turbines using sophisticated ML algorithms. Future research will include incorporating meteorological parameters as explanatory variables and applying advanced mathematical modeling techniques to create dynamic algorithms that enhance Wind Power Forecasting (WPF) through wavelet transformations. Additionally, we envision developing statistical charts and ML-driven dynamic models for fault detection in wind turbines. In the following research phase, we aim to improve wind energy forecasts by increasing the number of control variables to include a broader range of meteorological conditions such as temperature, humidity, wind speed, and air pressure. Furthermore, adaptive algorithms employing wavelet transformations will be integrated better to identify transient behaviors and fluctuations in wind patterns. This research aims to improve the reliability of wind energy forecasts and extend the use of wind energy. The recommendations and methods described in this project for increasing the effectiveness and credibility of wind energy contribute to the need to minimize the reliance on non-renewable energy sources. Furthermore, the proposed systems will adopt Machine learning and control charts to track the turbines' performance, enhancing efficiency and minimizing cost and cost breakdowns.

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

Table 7: The hyperparameter tuning technique for models is statistically analyzed.

	RF	XGB	GB	SVR	KN	Linear Regression
Number of values	10	10	10	10	10	10
Minimum	0.06	0.065	0.11	0.099	0.056	0.118
25% Percentile	0.069	0.07	0.13	0.1	0.06	0.1605
Median	0.069	0.07	0.13	0.1	0.06	0.168
75% Percentile	0.069	0.07	0.132	0.1038	0.061	0.168
Maximum	0.0719	0.071	0.14	0.15	0.072	0.2468
Range	0.0119	0.006	0.03	0.051	0.016	0.1288
10% Percentile	0.0609	0.0655	0.112	0.0991	0.0564	0.12
90% Percentile	0.07161	0.0709	0.1398	0.1465	0.0712	0.2389
95% CI of median						
Actual confidence level	97.85%	97.85%	97.85%	97.85%	97.85%	97.85%
Lower confidence limit	0.069	0.07	0.13	0.1	0.06	0.138
Upper confidence limit	0.069	0.07	0.138	0.115	0.064	0.168
Mean	0.06839	0.0696	0.1298	0.1064	0.0612	0.1679
Std. Deviation	0.003086	0.001647	0.007913	0.01604	0.004237	0.03268
Std. Error of Mean	0.0009758	0.0005207	0.002502	0.005073	0.00134	0.01033
Lower 95% CI of mean	0.06618	0.06842	0.1241	0.09492	0.05817	0.1445
Upper 95% CI of mean	0.0706	0.07078	0.1355	0.1179	0.06423	0.1913
Coefficient of variation	4.512%	2.366%	6.097%	15.08%	6.924%	19.46%
Geometric mean	0.06832	0.06958	0.1296	0.1055	0.06108	0.1652
Geometric SD factor	1.049	1.025	1.066	1.14	1.068	1.203
Lower 95% CI of geo. mean	0.06604	0.06838	0.1238	0.09603	0.05827	0.1447
Upper 95% CI of geo. mean	0.07068	0.0708	0.1356	0.1159	0.06402	0.1887
Harmonic mean	0.06825	0.06956	0.1293			

Author Contributions

Conceptualization, E.E; M.A; Data Collection, M.A; Analysis and Interpretation of results, M.A; I.I; Manuscript Preparation, I.I; E.E and M.A project administration, E.E. review & editing H.M.

Competing Interests

The authors declare no competing interests.

Data Availability Statement

The data that support the findings of this study are openly available at [`https://www.kaggle.com/datasetslook/berkerisen/wind-turbine-scada-dataset`].

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

DOI: https://doi.org/10.54216/FPA.190205
Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

Additional Information

Correspondence and requests for materials should be addressed to Corresponding author.
Reprints and permissions information Not applicable.
Ethical approval and consent to participate Not applicable.
Consent to Publish Not applicable.
Competing interests: the authors declare no competing interests.

References

[1] C. Haoran, Y. Xia, W. Wei, Z. Yongzhi, Z. Bo, and Z. Leiqi, "Safety and efficiency problems of hydrogen production from alkaline water electrolyzers driven by renewable energy sources," *International Journal of Hydrogen Energy*, vol. 54, pp. 700–712, 2024.

[2] A. Olabi, K. Obaideen, M. A. Abdelkareem, *et al.*, "Wind energy contribution to the sustainable development goals: Case study on london array," *Sustainability*, vol. 15, no. 5, p. 4641, 2023.

[3] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, "The role of renewable energy in the global energy transformation," *Energy Strategy Reviews*, vol. 24, pp. 38–50, 2019, ISSN: 2211-467X. DOI: `https://doi.org/10.1016/j.esr.2019.01.006`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S2211467X19300082`.

[4] W. Strielkowski, L. Civín, E. Tarkhanova, M. Tvaronavičienė, and Y. Petrenko, "Renewable energy in the sustainable development of electrical power sector: A review," *Energies*, vol. 14, no. 24, p. 8240, 2021.

[5] H. Alkabbani, A. Ahmadian, Q. Zhu, and A. Elkamel, "Machine learning and metaheuristic methods for renewable power forecasting: A recent review," *Frontiers in Chemical Engineering*, vol. 3, p. 665 415, 2021.

[6] L. Yin and Y. Sun, "Bilstm-inceptionv3-transformer-fully-connected model for short-term wind power forecasting," *Energy Conversion and Management*, vol. 321, p. 119 094, 2024.

[7] W. Huang, Y. Yin, H. Li, A. Xie, and Y. Fan, "High-speed railway express delivery volume forecast based on data-driven ensemble forecast approaches: The china case," *Expert Systems with Applications*, vol. 258, p. 125 196, 2024.

[8] S. Wang, S. Yan, H. Li, *et al.*, "Short-term prediction of photovoltaic power based on quadratic decomposition and residual correction," *Electric Power Systems Research*, vol. 236, p. 110 968, 2024.

[9] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," *Neural computation*, vol. 31, no. 7, pp. 1235–1270, 2019.

[10] Y. Wang, Y. Yu, S. Cao, X. Zhang, and S. Gao, "A review of applications of artificial intelligent algorithms in wind farms," *Artificial Intelligence Review*, vol. 53, no. 5, pp. 3447–3500, 2020.

[11] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, "A data-driven multi-model methodology with deep feature selection for short-term wind forecasting," *Applied Energy*, vol. 190, pp. 1245–1257, 2017.

[12] L. Landberg, "Short-term prediction of the power production from wind farms," *Journal of wind engineering and industrial Aerodynamics*, vol. 80, no. 1-2, pp. 207–220, 1999.

[13] J. R. Andrade and R. J. Bessa, "Improving renewable energy forecasting with a grid of numerical weather predictions," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 4, pp. 1571–1580, 2017.

[14] F. Baer, "Numerical weather prediction," in *Advances in Computers*, vol. 52, Elsevier, 2000, pp. 91–157.

[15] S. Aslam, H. Herodotou, S. M. Mohsin, N. Javaid, N. Ashraf, and S. Aslam, "A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids," *Renewable and Sustainable Energy Reviews*, vol. 144, p. 110 992, 2021.

[16] R. Meenal, D. Binu, K. Ramya, *et al.*, "Weather forecasting for renewable energy system: A review," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2875–2891, 2022.

DOI: https://doi.org/10.54216/FPA.190205
Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

[17] A. Mosavi, M. Salimi, S. Faizollahzadeh Ardabili, T. Rabczuk, S. Shamshirband, and A. R. Varkonyi-Koczy, "State of the art of machine learning models in energy systems, a systematic review," *Energies*, vol. 12, no. 7, p. 1301, 2019.

[18] G. Santamaría-Bonfil, A. Reyes-Ballesteros, and C. Gershenson, "Wind speed forecasting for wind farms: A method based on support vector regression," *Renewable Energy*, vol. 85, pp. 790–809, 2016.

[19] X.-w. Mi, H. Liu, and Y.-f. Li, "Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm," *Energy Conversion and Management*, vol. 151, pp. 709–722, 2017.

[20] K. Sareen, B. K. Panigrahi, T. Shikhola, and R. Sharma, "An imputation and decomposition algorithms based integrated approach with bidirectional lstm neural network for wind speed prediction," *Energy*, vol. 278, p. 127 799, 2023.

[21] A. A. Abdoos, H. Abdoos, J. Kazemitabar, M. M. Mobashsher, and H. Khaloo, "An intelligent hybrid method based on monte carlo simulation for short-term probabilistic wind power prediction," *Energy*, vol. 278, p. 127 914, 2023.

[22] L. Ye, Y. Li, M. Pei, Y. Zhao, Z. Li, and P. Lu, "A novel integrated method for short-term wind power forecasting based on fluctuation clustering and history matching," *Applied Energy*, vol. 327, p. 120 131, 2022.

[23] Y. Zhao, N. Guo, W. Chen, *et al.*, "Multi-step ahead forecasting for electric power load using an ensemble model," *Expert Systems with Applications*, vol. 211, p. 118 649, 2023.

[24] S. Liao, X. Tian, B. Liu, T. Liu, H. Su, and B. Zhou, "Short-term wind power prediction based on lightgbm and meteorological reanalysis," *Energies*, vol. 15, no. 17, p. 6287, 2022.

[25] M.-D. Liu, L. Ding, and Y.-L. Bai, "Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the arima to wind speed prediction," *Energy Conversion and Management*, vol. 233, p. 113 917, 2021.

[26] H. S. Dhiman, D. Deb, S. Muyeen, and A. Abraham, "Machine intelligent forecasting based penalty cost minimization in hybrid wind-battery farms," *International Transactions on Electrical Energy Systems*, vol. 31, no. 9, e13010, 2021.

[27] H. S. Dhiman, D. Deb, S. Muyeen, and I. Kamwa, "Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines," *IEEE Transactions on Energy Conversion*, vol. 36, no. 4, pp. 3462–3469, 2021.

[28] H. S. Dhiman, D. Deb, and J. M. Guerrero, "Hybrid machine intelligent svr variants for wind forecasting and ramp events," *Renewable and Sustainable Energy Reviews*, vol. 108, pp. 369–379, 2019.

[29] A. Zameer, J. Arshad, A. Khan, and M. A. Z. Raja, "Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks," *Energy conversion and management*, vol. 134, pp. 361–372, 2017.

[30] C. Li, S. Lin, F. Xu, D. Liu, and J. Liu, "Short-term wind power prediction based on data mining technology and improved support vector machine method: A case study in northwest china," *Journal of Cleaner Production*, vol. 205, pp. 909–922, 2018.

[31] T. Mahmoud, Z. Dong, and J. Ma, "An advanced approach for optimal wind power generation prediction intervals by using self-adaptive evolutionary extreme learning machine," *Renewable energy*, vol. 126, pp. 254–269, 2018.

[32] S. M. Malakouti, "Use machine learning algorithms to predict turbine power generation to replace renewable energy with fossil fuels," *Energy Exploration & Exploitation*, vol. 41, no. 2, pp. 836–857, 2023.

[33] H. Rashid, W. Haider, and C. Batunlu, "Forecasting of wind turbine output power using machine learning," in *2020 10th International conference on advanced computer information technologies (ACIT)*, IEEE, 2020, pp. 396–399.

[34] A. E. Gürel, Ü. Ağbulut, and Y. Biçen, "Assessment of machine learning, time series, response surface methodology and empirical models in prediction of global solar radiation," *Journal of Cleaner Production*, vol. 277, p. 122 353, 2020.

[35] U. Singh and M. Rizwan, "Enhancing wind power forecasting from meteorological parameters using machine learning models," *Journal of Renewable and Sustainable Energy*, vol. 14, no. 6, 2022.

[36] S. M. Malakouti, "Estimating the output power and wind speed with ml methods: A case study in texas," *Case Studies in Chemical and Environmental Engineering*, vol. 7, p. 100 324, 2023.

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025

[37] Q. Quan, Z. Hao, H. Xifeng, and L. Jingchun, "Research on water temperature prediction based on improved support vector regression," *Neural Computing and Applications*, pp. 1–10, 2022.

[38] E. Güvenç, G. Çetin, and H. Koçak, "Comparison of knn and dnn classifiers performance in predicting mobile phone price ranges," *Advances in Artificial Intelligence Research*, vol. 1, no. 1, pp. 19–28, 2021.

[39] A. K. Gupta, V. Singh, P. Mathur, and C. M. Travieso-Gonzalez, "Prediction of covid-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in indian scenario," *Journal of Interdisciplinary Mathematics*, vol. 24, no. 1, pp. 89–108, 2021.

DOI: https://doi.org/10.54216/FPA.190205

Received: December 24, 2024 Revised: February 07, 2025 Accepted: March 03, 2025