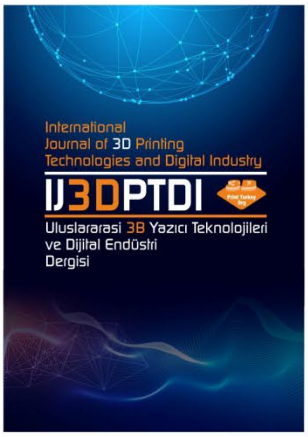

ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ VE DİJİTAL ENDÜSTRİ DERGİSİ

INTERNATIONAL JOURNAL OF 3D PRINTING TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 (Online)

URL: https://dergipark.org.tr/ij3dptdi

WIND TURBINE POWER PREDICTION USING MACHINE LEARNING MODELS: A CASE STUDY WITH REAL FARM DATA

Yazarlar (Authors): Abdullah Fatih Hocu, Gul Fatma Turker*

Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article): *Hocu A.F., Turker G. F., "Wind Turbine Power Prediction Using Machine Learning Models: A Case Study with Real Farm Data" Int. J. of 3D Printing Tech. Dig. Ind., 9(2): 395-404, (2025).*

DOI: 10.46519/ij3dptdi.1629937

Araştırma Makale/ Research Article

Erişim Linki: (To link to this article): https://dergipark.org.tr/en/pub/ij3dptdi/archive

WIND TURBINE POWER PREDICTION USING MACHINE LEARNING MODELS: A CASE STUDY WITH REAL FARM DATA

Abdullah Fatih Hocu[a], Gul Fatma Turker[a]*

[a]Suleyman Demirel University, Engineering and Natural Science Faculty, Computer Engineering Department, TURKEY

* *Corresponding Author: gulturker@sdu.edu.tr*

(**Received:** 30.01.25; **Revised:** 14.04.25; **Accepted:** 29.07.25)

ABSTRACT

The power generated from wind turbines is of critical importance as one of the fundamental components of sustainable and renewable energy systems. However, the complex and nonlinear nature of wind flow and the influence of interconnected factors make turbine power estimation significantly difficult. This study aims to evaluate the performance of different forecasting models using real-time data obtained from wind turbines and to determine the most effective model for wind power generation. The analyses are performed based on performance metrics that measure the agreement between the predicted and actual values. The study results reveal that the Decision Tree Regressor model provides the highest accuracy with 0.998 R^2 value and low error rates (RMSE: 0.151, MAE: 0.036) and that tree-based models are more effective in wind power estimation. These models, trained using large datasets, offer significant potential in terms of increasing power grid stability and ensuring the optimization of wind farms. The study shows that advanced methods used in turbine power estimation are an effective tool for optimizing renewable energy use by contributing to sustainable energy targets.

Keywords: Artificial Intelligence, Bigdata, Prediction, Renewable Energy, Sustainability

1. INTRODUCTION

Renewable energy has become a cornerstone of global efforts to combat climate change and reduce dependence on fossil fuels. Among these energy sources, wind energy has attracted attention due to its scalability and potential to generate significant amounts of energy without carbon emissions [1]. However, the inherent characteristics of wind, such as variability and discontinuity, make it difficult to accurately predict energy outputs, which creates critical challenges for integration into energy grids and system stability [2].

To address these challenges, various forecasting methods have been developed over the years. Traditional methods have been based on physical models and statistical analyses. For example, methods such as Autoregressive Integrated Moving Average (ARIMA) and Kalman filtering have tried to predict wind power based on historical data [3]. Numerical Weather Prediction (NWP) based physical models have improved forecast accuracy by simulating the effects of meteorological conditions [2]. Nonetheless, these conventional models fall short in capturing the complex and nonlinear behavior of wind dynamics [4].

In light of these limitations, recent years have witnessed a shift towards more data-driven approaches. The advent of Machine Learning (ML) and Deep Learning (DL) has ushered in a new era in wind power forecasting. Deep learning algorithms such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) have shown superior performance in capturing temporal and spatial dependencies in wind turbine data [5]. Hybrid approaches have improved prediction accuracy, especially with techniques such as the combination of Discrete Wavelet Transform (DWT) and LSTM [6]. Recently, meta-heuristic methods such as Sparrow Search Algorithm (SSA) have been used in wind power forecasts and have significantly improved the forecast accuracy [7]. In addition, hybrid ARIMA-LSTM models have shown high performance by capturing short- and long-term dependencies [8].

Despite these advancements, some challenges remain. The limited availability of high-quality data and the difficulty of generalizing models across different geographical regions limit the applicability of wind power forecasts [9]. Furthermore, errors in the accuracy of NWP models lead to performance degradation in short-term forecasts [10]. To mitigate these issues, hybrid strategies combining learning-based models with optimization techniques have been proposed. For example, performance improvement has been achieved by combining Teaching Learning Based Optimization (TLBO) algorithms with deep learning models [2]. In addition, Variational Mode Decomposition (VMD) and Gated Recurrent Unit (GRU) based models offer higher accuracy, especially in short-term forecasts [11].

Another important approach to improving forecasting performance is the effective use of turbine-level data. The utilization of turbine-level data is critical in improving forecast accuracy. These approaches have been improved by the development of hierarchical models that effectively utilize turbine-level information [12]. For example, new methods utilizing turbine-level information have led to improvements in probabilistic wind power forecasts [6,13].

In summary, while substantial progress has been made in the field of wind power forecasting, several persistent challenges remain to be addressed. Achieving high forecasting accuracy is essential not only for ensuring grid stability but also for the efficient integration of renewable energy into the power system. However, limitations such as the low quality and availability of data, regional and climatic variability, and the inherent constraints of traditional prediction models continue to hinder progress in this domain. The literature highlights that overcoming these challenges requires the development of more robust, generalizable, and adaptive approaches. Therefore, there is a pressing need for novel methods that can enhance model performance, accommodate diverse operating environments, and contribute to the reliable use of wind energy in sustainable power systems. In this study, the performance of various machine learning algorithms are compared using real-time data at the turbine level and an effective prediction model is developed. In the second section, information about the comprehensive dataset is given, and the details of the machine learning algorithms applied after pre-processing steps, along with evaluation parameters, are presented. In the third section, experimental results are presented and the models are compared. In the results section, the accuracy of the developed forecasting model and its contribution to power grid stability are discussed. In addition, the limitations of the study are evaluated and suggestions for future studies are presented.

2. MATERIAL AND METHODS

2.1. Data Source

In this study, real-time data were obtained from a wind farm located in the Denizli region of Turkey, where the energy production values of three turbines, each with three blades and a capacity of 3.4 MW (3400 kWp), were monitored for one year. The site has an elevation of 1618 meters and a hub height of 79.5 m. The dataset, comprising 526,221 rows and 22 columns, was collected between January 1 and December 31, 2024, with minute-level sampling recorded daily from an actual wind turbine in the region. These data enables the estimation of wind energy production over time. These dataset columns contain various measurements that track in detail the performance and operating parameters of a wind turbine. Parameters such as time data, device identifier, active power, wind speed, nacelle orientation, operating status, current and voltage values, frequency, reactive power, fault codes, rotor and generator rotational speeds, ambient temperature, hydraulic system pressure and total amount of energy supplied to the grid are used to monitor the instantaneous and cumulative performance, electrical characteristics and operating conditions of the turbine.

The dataset used in this study was obtained in real-time from an industrial-scale wind turbine site. It encompasses electrical, mechanical, and environmental variables measured under actual operational conditions. This enhances both the practical relevance of the study and its contribution to sustainable energy systems. Table 1 provides detailed descriptions of the dataset variables, including operational, environmental, and electrical parameters collected from wind turbines.

Table 1. Description of the wind turbine dataset variables.

Data Column	Description
time	Timestamp of the record (UTC)
device_id	Unique identifier assigned to each wind turbine
active_power	Active power output of the turbine (kW)
wind_speed	Wind speed at the turbine location (m/s)
yaw_direction	Nacelle (turbine head) direction (degrees)
wt_operationstate	Operational status/state of the turbine
current_v	Phase V current value (Amps)
current_u	Phase U current value (Amps)
current_w	Phase W current value (Amps)
voltage_v	Phase V voltage value (Volts)
voltage_u	Phase U voltage value (Volts)
voltage_w	Phase W voltage value (Volts)
frequency	Frequency of the generated electricity (Hz)
reactive_power	Reactive power output of the turbine (VAR)
error_code	Error codes, if any, detected during operation
rotor_rpm	Rotor rotational speed (Revolutions Per Minute)
generator_rpm	Generator rotational speed (RPM)
ambie_tmp	Ambient temperature near the turbine (°C)
hyd_press	Hydraulic system pressure
active_energy_export	Total energy exported to the electrical grid
eac	Active energy difference between consecutive measurements

The dataset used in the prediction algorithm includes electrical, mechanical, and environmental variables collected at the turbine level. Among these, electrical parameters-particularly current and voltage-played a decisive role in prediction performance, significantly enhancing model accuracy by providing strong predictive power.

2.2. Data Preprocessing

The performance and accuracy of wind turbine forecasting models can be significantly affected by missing or inaccurate data in the dataset. Such errors are usually caused by missing values, inconsistent data entries or irregularities in the dataset. In this study, a comprehensive data preprocessing process was implemented to improve the quality of the dataset and ensure the reliability of the prediction models.

Firstly, K-Nearest Neighbours (KNN), Linear Interpolation and Mode Interpolation methods were used to fill the missing data. These methods provided a statistically consistent filling of missing values. In addition, redundant and repetitive columns (redundant features) in the dataset were identified and removed. In order to create a data structure suitable for time series analysis, new features derived from the 'time' column were created. These features include 'hour', 'day', 'month', 'year', 'day name'. These inferences allowed the model to learn temporal dynamics better.

In order to increase the generalization capability of the dataset and to prevent overfitting of the model, cross-validation method was applied. In addition, normalization was performed to eliminate the scale differences between the features and to increase the convergence speed of the model. These steps made the dataset suitable for machine learning models. In addition, statistical analyses were performed on input features and power. This process contributed to eliminating inconsistencies in the dataset and increasing model accuracy.

Feature Engineering: Feature extraction is the process of creating numerical features from raw data that can be used in machine learning models. The number of input features can be very large and many of them may have low correlation with the target variables. Feature selection methods are a critical step to improve model performance, reduce training time and improve the interpretability of models [14].

In this study, new features are derived to better model wind turbine performance and improve prediction accuracy. These features include energy exchange ratio (active energy / reactive energy), rotor and generator speed ratio, current unbalance (standard deviation of currents), voltage unbalance (standard deviation of voltages) and power efficiency (active power / wind speed). These derived characteristics better reflect the dynamic behavior of the system, strengthening the predictive capability of the model and enabling a more effective analysis of critical parameters related to power generation.

2.3. Machine Learning

Within the scope of machine learning, various regression-based machine learning models are applied in this study for the prediction of wind turbine data. These models aim to estimate the dependent variable based on independent features using learning algorithms [15]. In the study, nine machine learning models were trained according to the characteristics of the dataset and the results were analyzed. The data were divided into 80% for training the model and 20% for testing.

Among the models, Decision Tree Regressor, based on decision trees, stands out with its capacity to make fast and effective predictions by branching the data [14]. XGBoost Regressor, a more advanced algorithm, provides high accuracy and performance by using the gradient boosting method [16]. Similarly, Gradient Boosting Regressor aims to reduce the error by successively optimizing a set of weak predictors.

Furthermore, the Extra Tree Regressor model offers greater generalization capacity by using randomized decision tree principles [14,17]. K-Neighbors Regressor offers a simple and efficient approach, basing predictions on the values of the KNN. Linear Regression, a more basic model, makes predictions assuming a linear relationship between the target and independent variables. However, this model may have limitations on non-linear datasets. In addition, more sophisticated models such as Ridge, Elastic Net and Lasso try to avoid overfitting by using regularization techniques [18-20]. This process has enabled the identification of the most suitable algorithms for wind turbine forecasting models.

2.4. Performance Metrics of the Model

Error metrics used to evaluate the performance of machine learning algorithms play a critical role in measuring how well the model fits with real values and its generalization capability. These metrics help to evaluate the effectiveness and reliability of the model by determining its predictive accuracy and predictive power.

Mean Absolute Error (MAE) is a metric used to determine how close the predictions are to the true values and evaluates the average error of the predictions. This metric is calculated by Eq. 1 [21].

$$MAE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left|Z_j - \hat{Z}_j\right|} \quad (1)$$

Root Mean Square Error (RMSE) measures the magnitude of deviations of the model's predictions from the true values; it indicates how well the predicted values agree with the true observations and is a metric reflecting the error rate. To compare the prediction accuracy of different models, RMSE is calculated by Eq. 2 [21-22].

$$\text{RMSE} \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(Z_j - \hat{Z}_j\right)^2} \quad (2)$$

R-squared (R^2) is a measure of how well a model fits the data; high values indicate the explanatory power of the model. Eq. 3 [23].

$$R^2 = 1 - \frac{\sum_{j=1}^{n}\left(z_j - \hat{z}_j\right)^2}{\sum_{j=1}^{n}\left(z_j - \hat{z}_j\right)} \quad (3)$$

Mean Squared Error (MSE) is a widely used evaluation metric that quantifies the average of the squared differences between predicted and actual values. It provides a measure of how close the regression model's predictions are to the true outcomes. The MSE is calculated by Eq. 4.

$$MSE = \frac{1}{m}\sum_{j=1}^{m}\left(Z_j - \hat{Z}_j\right)^2 \quad (4)$$

3. EXPERIMENTAL FINDINGS

In this study, the power generation performance and operating dynamics of the wind turbines are analyzed in detail using annual real-time data obtained from the wind farm. The analyses of the monthly energy production values reveal that the turbines provide a high stability in energy production throughout the year, and these findings are presented in detail in Figure 1. From January to December, the total energy production ranged between 2.5×10^{12} kWh and 3.0×10^{12} kWh, with a slight increase in production in the last quarter of the year (October, November, December). This increase can be attributed to the increase in wind speeds during this period of the year. However, the generation remained relatively constant during the summer months (June, July, August), indicating that the system operates continuously and efficiently despite the seasonal variations in wind speed. The results show that wind turbines provide a reliable contribution to the energy needs in the region.

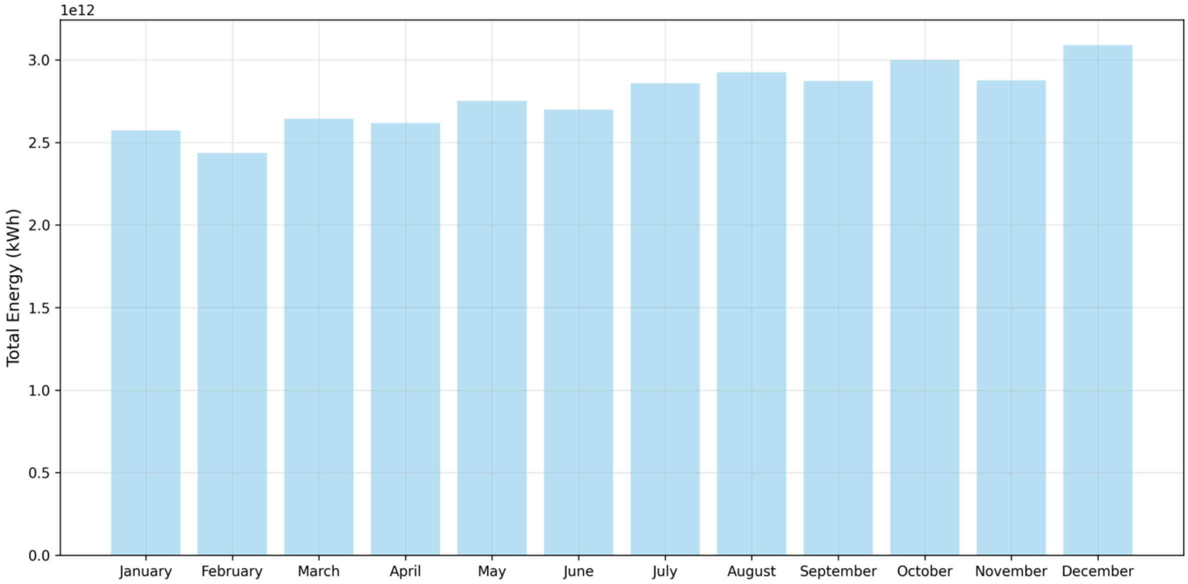

Figure 1. Total montly energy production.

Figure 2 illustrates the importance of various input features in predicting active power output, based on the linear regression coefficients. The results clearly underscore the dominant role of electrical current variables in the model's predictive performance. Among all variables, three-phase current values (current_w, current_v, current_u) exhibited the highest regression coefficients. While current_w and current_v contributed positively, current_u presented a strong negative coefficient, likely reflecting phase-specific imbalances in the system.

The differences between the phases are due to asymmetric loading and instantaneous power regulation in the system. In real wind turbine data, the contribution of each phase to the active power is not equal due to factors such as rotor position, wind direction and inverter operating characteristics. The higher effect observed especially in the W phase can be associated with the situations where this phase carries more current under load or that phase transmits more energy at the inverter output. Such imbalances are commonly observed in field operations.

Voltage components (voltage_v and voltage_w) also had a moderate yet positive influence on the model output, indicating their supplementary role in power estimation. Conversely, mechanical indicators such as wind speed and rotor RPM, although theoretically significant and highly correlated with active power, showed relatively lower predictive power within the linear regression model. This suggests that for short-term forecasting, electrical signals provide more direct and immediate predictive value than mechanical inputs.

Environmental parameters such as ambient temperature and hydraulic pressure contributed minimally, offering only limited explanatory power related to operational efficiency. Similarly, frequency and reactive power had negligible influence, indicating their minor role in determining active power within this context.

Overall, the integration of highly impactful electrical variables, particularly current and voltage, substantially enhanced the model's prediction accuracy. These findings reinforce the superiority of electrical parameters over mechanical and environmental variables for short-term active power forecasting in wind energy systems.

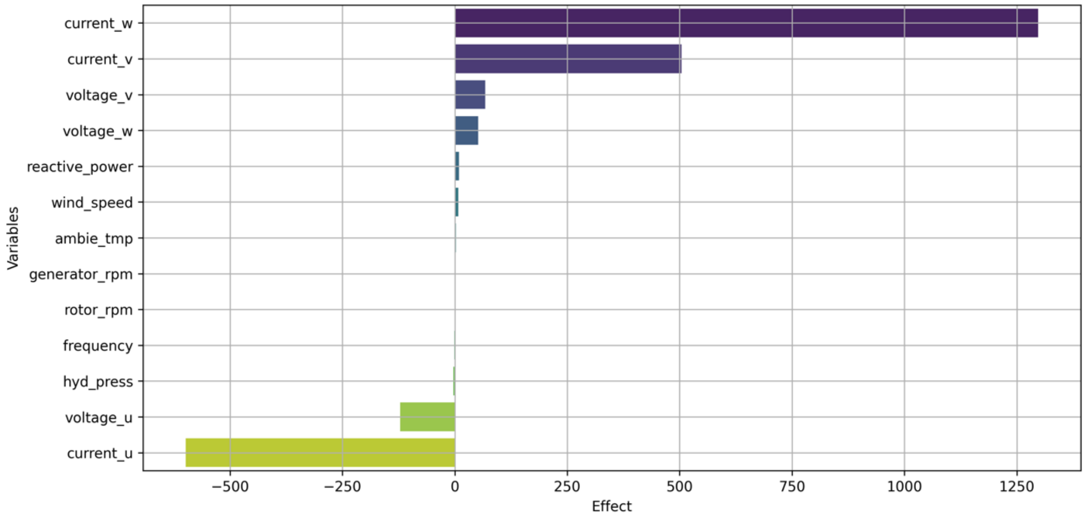

Figure 2. The impact of variables on active power based on linear regression coefficients.

The correlation matrix illustrates the interdependencies among various wind turbine parameters. A very strong positive correlation was observed between active power and current in phases U, V, and W (r = 1.0), indicating that electrical current directly determines active power generation. Additionally, rotor RPM and generator RPM also showed a strong correlation with active power (r = 0.83), underlining the importance of mechanical rotation in energy conversion. Interestingly, voltage values (U, V, W) displayed weaker correlations with active power (ranging from 0.16 to 0.25), and frequency showed an even lower correlation (r = 0.042). These findings reinforce that mechanical dynamics such as rotor speed, along with electrical current, are the dominant factors in power production, whereas voltage and frequency play a comparatively limited role.

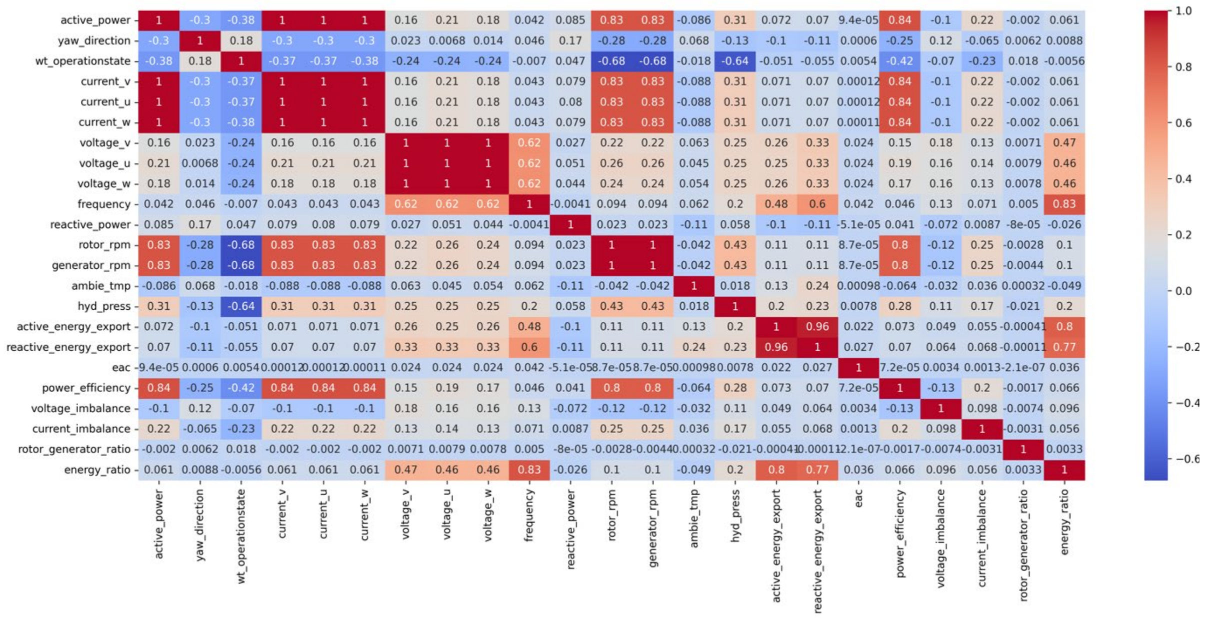

Figure 3. Wind farm parameter correlation matrix.

In this study, various machine learning models were applied to predict energy production using real-time farm data obtained from wind turbines and the performances of these models are comprehensively given in Table 2 with metrics such as accuracy (R^2), error rates (RMSE, MAE).

It compares the performance of various machine learning algorithms in predicting wind turbine energy production. Decision Tree Regressor showed the highest performance with R^2 value of 0.998, RMSE 0.151 and MAE 0.036, and provided the most accurate predictions by providing a very good fit to the data. XGBoost Regressor and Extra Tree Regressor provided an accuracy close to Decision Tree with R^2 values of 0.995 and 0.989, respectively, and were among the reliable models with low error rates. Gradient Boosting Regressor provided an acceptable accuracy with R^2 value of 0.962, but was considered a less effective model due to higher RMSE and MAE values. K-Neighbors Regressor showed a moderate performance with R^2 value of 0.912. However, Linear Regression and Ridge models had low accuracy with R^2 values of 0.806, and their high error rates showed that they could not capture the data complexity sufficiently. Elastic Net and Lasso models had the lowest accuracy with R^2 values of 0.734 and 0.729, and they exhibited poor performance with high error rates.

Table 2. Machine learning regression model results for wind power prediction.

ML Regression Algorithms	R_Squared	RMSE	MAE
DecisionTree	0.998	0.151	0.036
XGBoost	0.995	0.256	0.114
ExtraTree	0.989	0.390	0.086
GradientBoosting	0.962	0.740	0.433
Kneighbours	0.912	1.126	0.746
Linear	0.806	1.679	1.123
Ridge	0.806	1.679	1.123
ElasticNet	0.734	1.965	1.235
Lasso	0.729	1.983	1.235

Figure 4 compares the agreement between predicted and actual values of different machine learning models. The analysis revealed that tree-based models provide the highest accuracy in energy production estimation by better capturing complex data relationships. While these models provide effective and reliable results in wind energy estimation, it was observed that simple models cannot adequately represent the data complexity and therefore exhibit lower performance. This clearly shows that complex and optimized algorithms provide more reliable results in wind power estimation.

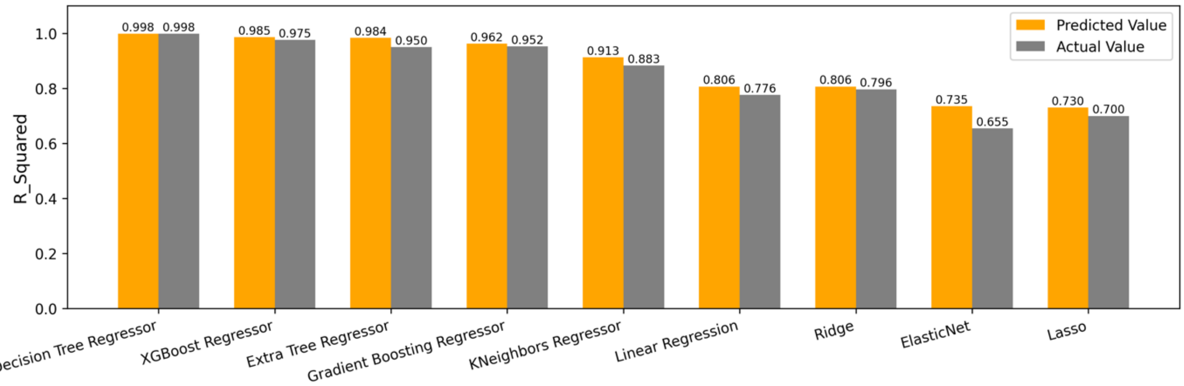

Figure 4. Predicted and actual values of the machine learning models of R^2 score.

The accuracy of the models developed within the scope of the study was tested not only on the turbine data used for training, but also on the data of a different turbine located in the same region. This validation process is critical to evaluate the generalization capacity of the models and to determine the consistency of performance on new data. The results revealed that the Decision Tree Regressor model provided a high fit ($R^2 = 0.998$) in both the training and validation phases. This finding proves that the model is not limited to a specific turbine, but can also make effective predictions with data obtained from different turbines in the same region.

This approach emphasizes a model development process that ensures consistency within regional differences and increases the potential for the model to be used in real-world applications. It also demonstrates the critical importance of region-specific turbine data on forecast performance. In the future, performing similar validation processes on turbines in different regions can further strengthen the generalization capacity of the model and provide a solid foundation for wider applications. In this context, it is aimed to provide more reliable and sustainable solutions in renewable energy systems.

3.1. Discussion

This study evaluates the effectiveness of machine learning algorithms in wind power forecasting using real-time data from wind turbines. The analysis revealed that the Decision Tree Regressor model achieved superior performance with high accuracy ($R^2 = 0.998$), effectively modeling complex and nonlinear relationships. These findings are consistent with previous research, which highlights the effectiveness of tree-based models in wind energy forecasting.

Traditional physical and statistical methods often fall short in modelling regional and meteorological variability. In contrast, machine learning and artificial intelligence-based approaches offer strong potential to address these challenges. Particularly, hybrid models and transfer learning techniques have shown notable improvements in forecast accuracy [24]. However, this study demonstrates that even with simple models and high-quality turbine-level data, competitive results can be achieved.

Some limitations should be considered. The dataset covers only a specific region and time frame, which may restrict the generalization of the findings to other geographic and climatic conditions. The quality and representativeness of turbine data significantly influence prediction accuracy. Therefore, the use of larger and more diverse datasets, along with adaptation strategies such as transfer learning, will be critical for enhancing the generalization capacity of forecasting models.

Future research could focus on the integration of hybrid model architectures and meta-heuristic optimization techniques to improve robustness and scalability. Additionally, the development of explainable forecasting systems, such as transformer-based models enhanced with attention mechanisms, may enhance both the transparency and usability of prediction tools. Ultimately, such innovative approaches can support the reliable integration of renewable energy into power grids and contribute to the achievement of sustainable energy goals.

4. RESULTS

This study investigates the applicability of machine learning algorithms for wind power forecasting using real-time data from wind turbines. Among the tested models, the Decision Tree Regressor yielded the highest accuracy ($R^2 = 0.998$) and the lowest error metrics (RMSE: 0.151, MAE: 0.036), demonstrating its robustness in capturing complex data patterns. In contrast, linear models such as Linear Regression and Ridge showed limited performance, failing to adequately model the nonlinearities inherent in wind energy data. The findings highlight the suitability of tree-based models for wind power forecasting and their potential to enhance grid stability and support the sustainable integration of wind energy into power systems. By leveraging turbine-level data, machine learning approaches offer accurate and scalable solutions for modern energy management.

REFERENCES

1. Lu, H., Zhang, L., Tian, C., Niu, T., & Wei, W., "Volatility index prediction based on a hybrid deep learning system with multi-objective optimization and mode decomposition", Energy Conversion and Management, Vol. 323, Pages 119155, 2024.

2. Sulaiman, M. H., & Mustaffa, Z., "Enhancing wind power forecasting accuracy with hybrid deep learning and teaching-learning-based optimization", Cleaner Energy Systems, Vol. 9, Pages 100139, 2024.

3. Tian, L., & Wei, Z., "Integration of VMD and neuro-fuzzy systems for wind speed analysis", Energy Conversion and Management, Vol. 323, Pages 119155, 2025.

4. Mehmood, Z., & Wang, Z., "Wind turbine energy forecasting using real wind farm's measurement data and performance of gene expression programming analytical model in comparison to traditional algorithms", International Journal of Green Energy, Vol. 22, Issue 2, Pages 414–431, 2025.

5. Bashir, T., Wang, H., Tahir, M. F., & Zhang, Y., "Short-term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN", Renewable Energy, Vol. 239, Pages 122055, 2025.

6. Gilbert, A., & Huang, L., "Hierarchical approaches for turbine-level wind power forecasting", Renewable Energy, Vol. 167, Pages 119155, 2020.

7. Shao, L., Huang, W., Liu, H., & Li, J., "Study of wind power prediction in ELM based on improved SSA", IEEJ Transactions on Electrical and Electronic Engineering, 2025.

8. AlShafeey, M., & Csaki, C., "Adaptive machine learning for forecasting in wind energy", Heliyon, Vol. 10, Pages e34807, 2024.

9. Ukoba, K., Odebiyi, A., & Oghenevwogaga, J., "Harnessing machine learning for sustainable futures: Advancements in renewable energy and climate change mitigation", Bulletin of the National Research Centre, Vol. 48, Pages 99, 2024.

10. Yang, M., Guo, Y., Huang, T., & Zhang, W., "Power prediction considering NWP wind speed error tolerability", Applied Energy, Vol. 377, Pages 124720, 2025.

11. Liu, Z., Guo, H., Zhang, Y., & Zuo, Z., "A comprehensive review of wind power prediction based on machine learning", Energies, Vol. 18, Issue 350, 2025.

12. Mokarram, M., et al., "Adaptability of forecasting models across geographies", Sustainable Energy Technologies and Assessments, Vol. 54, Pages 104070, 2025.

13. Chen, X., Han, B., Wang, X., Zhao, J., Yang, W., & Yang, Z., "Machine learning methods in weather and climate applications: A survey", Applied Energy, Vol. 13, Pages 12019, 2021.

14. Ahmad, M. W., Reynolds, J., & Rezgui, Y., "Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees", Journal of Cleaner Production, Vol. 203, Pages 810-821, 2018.

15. Panda, S. K., & Mohanty, S. N., "Time series forecasting and modeling of food demand supply chain based on regressors analysis", IEEE Access, Vol. 11, Pages 42679-42700, 2023.

16. Velthoen, J., Dombry, C., Cai, J. J., & Engelke, S., "Gradient boosting for extreme quantile regression", Extremes, Vol. 26, Issue 4, Pages 639-667, 2023.

17. Geurts, P., Ernst, D., & Wehenkel, L., "Extremely randomized trees", Machine Learnig, Vol. 63, Pages 3-42, 2006.

18. He, H. J., Zhang, C., Bian, X., An, J., Wang, Y., Ou, X., & Kamruzzaman, M., "Improved prediction of vitamin C and reducing sugar content in sweetpotatoes using hyperspectral imaging and

LARS-enhanced LASSO variable selection", Journal of Food Composition and Analysis, Vol. 132, Pages 106350, 2024.

19. Ranstam, J., & Cook, J. A., "LASSO regression", Journal of British Surgery, Vol. 105, Issue 10, Pages 1348-1348, 2018.

20. Diebold, F. X., & Shin, M., "Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives", International Journal of Forecasting, Vol. 35, Issue 4, Pages 1679-1691, 2019.

21. Nayak, J., Vakula, K., Dinesh, P., Naik, B., & Pelusi, D., "Intelligent food processing: Journey from artificial neural network to deep learning", Computer Science Review, Vol. 38, Pages 100297, 2020.

22. Khan, M. I. H., Sablani, S. S., Nayak, R., & Gu, Y., "Machine learning-based modeling in food processing applications: State of the art", Comprehensive Reviews in Food Science and Food Safety, Vol. 21, Issue 2, Pages 1409-1438, 2022.

23. Chicco, D., Warrens, M. J., & Jurman, G., "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation", PeerJ Computer Science, Vol. 7, Pages e623, 2021.

24. Abdel-Aty, A.-H., et al., "Boosting wind turbine performance with advanced smart power prediction", Alexandria Engineering Journal, Vol. 96, Pages 58–71, 2024.