International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

Machine Learning-Based Active Wind Power Prediction Using Comparative Regression Models

Sai Bhanu Prasad[1], Pallavi Malloju[2]

[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

Abstract. The rapid expansion of renewable energy technologies has increased the need for accurate forecasting techniques that can effectively estimate power generation from wind energy systems. Because wind conditions fluctuate continuously due to changing atmospheric and environmental factors, predicting the amount of electrical power generated by wind turbines remains a challenging task. Reliable forecasting models are essential for improving grid stability, optimizing energy distribution, minimizing operational uncertainty, and supporting efficient utilization of renewable energy resources. In this research, a estimate active wind power by comparing the performance of multiple regression algorithms. The study employs six widely used supervised learning models, namely Linear Regression, Ridge Regression, Lasso Regression, K-Neighbors Regression, Decision Tree Regression, and Gradient Boosting Regression. A real-world wind turbine dataset obtained from the Kaggle repository, consisting of meteorological measurements and turbine operational parameters recorded at regular intervals, is used for model development and evaluation. Before training the models, the dataset undergoes comprehensive preprocessing, including missing value estimation through imputation, detection and treatment of abnormal observations, feature correlation analysis, K-Neighbors Regression algorithm consistently delivers superior prediction accuracy compared with the remaining regression techniques, producing the lowest prediction errors and the highest coefficient of determination. The outcomes of this study confirm that machine learning-based regression methods provide an effective solution for active wind power forecasting and can significantly contribute to intelligent energy management, improved scheduling of renewable power generation, and the reliable operation of modern smart grid systems.

keywords: Active Wind Power Prediction, Renewable Energy, Machine Learning, Linear Regression, Ridge Regression, Lasso Regression, K-Neighbors Regression, Decision Tree, Gradient Boosting, Principal Component Analysis, Wind Turbine, Energy Forecasting.

I. Introduction

Renewable energy resources have become increasingly important as countries seek environmentally sustainable alternatives to conventional fossil fuels. Among the various renewable energy sources, wind energy has emerged as one of the fastest-growing technologies because of its clean operation, unlimited availability, and minimal environmental impact. The widespread installation of wind farms across different regions has significantly contributed to reducing greenhouse gas emissions while supporting global energy demands. However, the intermittent nature of wind creates considerable challenges in maintaining a stable balance between electricity generation and consumption.

output plays a critical role in ensuring reliable grid operation, reducing operational costs, and improving energy scheduling. Since wind speed continuously varies according to atmospheric conditions, the generated electrical power also fluctuates significantly. These variations make precise forecasting essential for energy providers, allowing them to optimize power generation, reduce reserve requirements, and improve decision-making in smart grid environments.

provided efficient solutions for modeling complex nonlinear relationships between meteorological variables and power generation. Unlike traditional statistical forecasting techniques, machine learning algorithms automatically learn hidden patterns from historical data and provide highly accurate predictions without requiring explicit mathematical models. Regression-based learning methods have become particularly effective because they estimate continuous output values while handling multiple input variables simultaneously.

This research presents a comparative study of six machine learning regression algorithms, namely experimental framework utilizes operational wind turbine data collected from the Kaggle repository. Before model training, the dataset undergoes comprehensive preprocessing that includes missing value imputation, outlier removal, for practical wind energy forecasting applications.

II. Literature Survey

The rapid advancement of renewable energy technologies has encouraged extensive research on wind power forecasting using machine learning and artificial intelligence techniques. Traditional forecasting methods mainly relied on statistical approaches, which often struggled to model the nonlinear and uncertain behavior of wind characteristics. As computational intelligence evolved, machine learning algorithms became increasingly popular because of their ability to analyze complex relationships among meteorological variables and generate more accurate predictions.

Several researchers have investigated Support Vector Machines, Artificial Neural Networks, and ensemble learning techniques for forecasting wind speed and wind power. These approaches demonstrated significant improvements over conventional mathematical models by effectively learning nonlinear data distributions. Artificial Neural Networks have been widely adopted due to their capability to capture intricate patterns within historical wind datasets, while Support Vector Regression has shown promising performance in handling high-dimensional prediction problems.

Regression-based machine learning algorithms have also received considerable attention Linear Regression remains one of the simplest forecasting methods and serves as a baseline for performance comparison. Ridgestability. Decision Tree Regression partitions data into hierarchical decision structures, enabling effective prediction of nonlinear relationships. Gradient Boosting further enhances prediction performance by combining multiple weak learners into a powerful ensemble model. K-Neighbors Regression predicts output values by analyzing the similarity between neighboring observations, making it highly suitable for localized prediction tasks.

Recent studies have further incorporated dimensionality features while preserving the majority of data variance. Researchers have also emphasized the importance of handling missing values and outliers before model training to improve prediction reliability. Although numerous forecasting techniques have been proposed, selecting the most appropriate regression model remains dependent on dataset characteristics and feature quality. Motivated by these findings, this work performs a comparative evaluation of six regression algorithms using real-world wind turbine data to determine the most effective model for active wind power prediction.

III. Research Methodology

The proposed wind power forecasting framework follows a systematic machine learning workflow designed to maximize prediction accuracy while ensuring reliable model evaluation. Initially, operational wind turbine data are collected from a publicly available Kaggle dataset containing meteorological measurements, turbine operating parameters, and corresponding active power generation values. The dataset includes multiple environmental and mechanical gearbox temperature, ambient temperature, blade pitch angle, nacelle position, and reactive power, all of which influence wind energy production.

During the preprocessing stage, the dataset is carefully examined to identify missing values, inconsistent observations, and abnormal records. technique implemented through the Iterative Imputer, allowing incomplete records to be reconstructed using relationships among available attributes. Outliers that could negatively affect prediction accuracy are detected and replaced before further processing preserving approximately 99% of the original dataset variance while significantly reducing feature dimensionality.

Following preprocessing, exploratory data analysis is performed to understand relationships among variables through correlation analysis and statistical visualization. The processed dataset is then divided into training and testing subsets, enabling objective evaluation of model performance on previously unseen data.—are trained independently using identical datasets. Cross-validation and early stopping techniques are incorporated where appropriate to minimize overfitting and improve generalization capability.

Finally, the trained models predict active Comparative analysis of these evaluation metrics identifies the regression model that provides the highest prediction accuracy while maintaining computational efficiency. Experimental results demonstrate that the K-Neighbors Regression model consistently achieves superior forecasting performance compared with the remaining machine learning models, making it the most suitable approach for active wind power prediction within the selected experimental environment.

The forecasting framework evaluates six supervised machine learning regression algorithms, namely Linear Regression, Ridge Regression, Lasso Regression, K-Neighbors Regression, Decision Tree Regression, and Gradient Boosting Regression. Each model is trained independently using identical training data to ensure a fair comparison. Linear Regression establishes a baseline prediction model by learning by recursively partitioning the feature space, whereas Gradient Boosting incrementally combines multiple weak learners to generate a stronger predictive model. K-Neighbors Regression estimates active wind power by analyzing the similarity between neighboring observations and predicting values based on the nearest data points.

To improve model reliability, appropriate validation techniques are incorporated throughout the training process to minimize overfitting and enhance generalization performance. After training is completed, all regression models are tested using the reserved testing dataset. Their predictive performance is quantitatively measured using overall model effectiveness.

A comparative analysis of the obtained results reveals that the K-Neighbors Regression algorithm delivers the highest prediction performance among all evaluated models. It achieves lower prediction errors and a higher R^2 Score than Based on these experimental findings, the proposed machine learning framework demonstrates that K-Neighbors Regression is the most effective approach for active wind power forecasting under the selected dataset and experimental configuration, making it a reliable solution for renewable energy prediction and intelligent power management systems.

IV. Existing System

To identify patients who may be at risk of developing high Existing wind power forecasting systems primarily depend on conventional statistical methods and machine learning regression techniques to estimate the amount of electrical power generated by wind turbines. These systems generally utilize historical operational data collected from wind farms, including meteorological measurements and turbine operating parameters such as wind speed, wind direction, rotor speed, generator speed, ambient temperature, gearbox temperature, blade pitch angles, nacelle position, and reactive power. Before model development, the collected data undergoes preprocessing procedures to handle missing values, eliminate inconsistencies, and improve data quality. commonly applied to remove redundant information and reduce computational complexity. are trained using historical observations to predict future active wind power generation. The performance of these forecasting models is generally assessed using regression evaluation metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2 Score), enabling researchers to compare prediction accuracy among different algorithms.

Although these existing forecasting approaches have demonstrated satisfactory performance, they still possess several limitations that affect prediction reliability under practical operating conditions. Linear Regression often struggles to model generation. Ridge and Lasso Regression improve model stability through regularization but may fail to capture intricate interactions among multiple turbine parameters. Decision Tree Regression is prone to overfitting when trained on noisy or highly variable datasets, while Gradient Boosting requires careful hyperparameter tuning and increased computational resources to achieve optimal performance. Furthermore, forecasting accuracy can be significantly influenced by missing data, outliers, rapidly changing weather conditions, seasonal variations, and the intermittent nature of wind energy. These challenges indicate that selecting an appropriate regression algorithm and applying effective data preprocessing techniques are essential for improving forecasting precision. Consequently, there remains a continuous need for robust machine learning frameworks capable of producing more accurate, stable, and computationally efficient wind power predictions for modern renewable energy management systems.

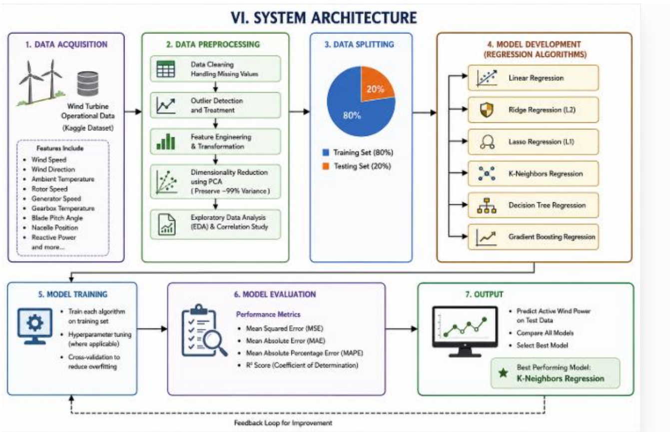
V. Proposed System

Accurately forecasting active wind power by utilizing multiple regression algorithms and a comprehensive data preprocessing pipeline. The system begins with the acquisition of a real-world wind turbine dataset obtained from the Kaggle repository, which contains meteorological information and turbine operational measurements recorded at regular time intervals. These records include important parameters such as wind speed, wind direction, ambient temperature, generator speed, rotor speed, gearbox temperature, blade pitch angles, nacelle position, reactive power, and other turbine-related attributes that directly influence wind energy production. Prior to model development, the dataset undergoes an extensive preprocessing stage where technique, abnormal observations are identified and treated, and inconsistent records are corrected to improve overall data quality. To eliminate redundant information and reduce computational complexity, Principal Component Analysis (PCA) is employed, preserving nearly all of the dataset's useful information while transforming it into a smaller and more efficient feature space. Exploratory Data Analysis (EDA) and correlation analysis are further performed to understand the relationships among variables and identify the most significant features contributing to active wind power generation.

International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

After preprocessing, the refined dataset is divided into training and testing subsets to ensure an unbiased evaluation of model performance. The proposed framework implements six supervised machine learning regression algorithms: Linear Regression, Ridge Regression, Lasso Regression, K-Neighbors Regression, Decision Tree Regression, and Gradient Boosting Regression. Each model is independently trained using identical input features, allowing a fair comparison of their prediction capabilities. Linear Regression serves as the baseline model for establishing linear relationships, while Ridge and Lasso Regression utilize regularization techniques to reduce overfitting and improve model stability. Decision Tree Regression learns nonlinear decision rules by recursively partitioning the feature space, whereas Gradient Boosting combines multiple weak decision trees to construct a more powerful predictive model. K-Neighbors Regression estimates active wind power by identifying the closest neighboring observations within the feature space and calculating predictions based on their similarity. Once the training process is completed, all models are evaluated using the testing dataset with standard regression performance metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2 Score). Comparative experimental analysis demonstrates that the K-Neighbors Regression model achieves the highest prediction accuracy and the lowest forecasting error among all evaluated regression algorithms. The proposed framework therefore provides a reliable, efficient, and scalable solution for active wind power forecasting, making it highly suitable for renewable energy management, smart grid operation, power scheduling, and sustainable electricity generation.

VI. System Architecture



The proposed system architecture provides a complete machine learning pipeline for forecasting active wind power using historical wind turbine operational data. The framework begins with the data acquisition stage, where a real-world dataset is collected from the Kaggle repository. The dataset contains measurements recorded at regular time intervals from wind turbines operating under different environmental conditions. It includes various meteorological and turbine-related parameters such as wind speed, wind direction, ambient temperature, rotor speed, generator speed, gearbox temperature, blade pitch angles, nacelle position, reactive power, and several other operational variables. These features collectively influence the amount of electrical energy generated by wind turbines and therefore serve as the primary input variables for the forecasting system.

Once the dataset is collected, it enters the data preprocessing stage, which is essential for improving data quality before model training. Initially, the dataset is inspected to detect) technique through the Iterative Imputer, enabling incomplete records to be reconstructed based on relationships among existing features. The system also performs outlier detection to identify unrealistic values that may negatively affect prediction accuracy. These abnormal observations are appropriately treated or replaced to ensure a clean and reliable dataset. Feature engineering techniques are then applied to transform the raw attributes into more informative representations. Furthermore, Principal Component Analysis (PCA) is utilized to reduce feature dimensionality while preserving approximately 99% of the original data variance. This process removes redundant information, minimizes multicollinearity, decreases computational complexity, and improves the learning efficiency of the regression algorithms. Exploratory Data Analysis (EDA) and correlation analysis are subsequently performed to understand the relationships among variables and identify the most influential factors affecting active wind power generation.

Following preprocessing, the refined dataset is divided into training and testing subsets using an 80:20 ratio. Approximately 80% of the observations are used to train the regression models, while the remaining 20% are reserved for performance evaluation. This separation ensures that the predictive capability of each model is assessed using previously unseen data, providing an unbiased estimate of its real-world forecasting performance.

The next stage focuses on model development, where six supervised machine learning regression algorithms are implemented for comparative analysis. These and Gradient Boosting Regression. Linear Regression establishes a mathematical relationship between input variables and active wind power through a linear equation. Ridge Regression introduces L2 regularization to reduce coefficient variance and minimize overfitting, whereas Lasso Regression applies L1 regularization to eliminate less significant features and simplify the prediction model. Decision Tree Regression recursively partitions the dataset into decision nodes to capture nonlinear relationships among variables. improve forecasting performance. K-Neighbors Regression predicts wind power by analyzing the similarity between neighboring observations in the feature space and estimating output values based on the nearest data points.

After model construction, the system proceeds to the training phase, during which each regression algorithm learns the underlying relationships between the input features and the target variable using the training dataset. Appropriate validation strategies and hyperparameter optimization are applied wherever necessary Once training is completed, the developed models are tested using the reserved testing dataset to evaluate their prediction capabilities under unseen conditions.

The effectiveness of every regression model is measured during the performance evaluation stage using standard regression metrics, including Mean Square]estimation error, and the overall reliability of each forecasting model. Lower values of MSE, MAE, and MAPE indicate better predictive performance, while a higher R^2 Score reflects a stronger relationship between predicted and actual wind power values.

Finally, the output stage generates the predicted active wind power for new input data and compares the performance of all implemented regression algorithms. Based on the experimental evaluation, the K-Neighbors Regression model demonstrates the highest forecasting accuracy and the lowest prediction error among all six machine learning techniques. Consequently, it is selected as proposed framework. The overall architecture provides an efficient, scalable, and intelligent forecasting solution that can assist renewable energy providers in optimizing wind farm operations, improving power scheduling, enhancing grid stability, reducing operational uncertainty, and supporting the reliable integration of wind energy into modern smart grid infrastructures.

VII. Conclusion

Using actual wind turbine operational data and various regression algorithms, this research presents an efficient machine learning framework for forecasting active wind output. Data collection, preprocessing, outlier handling, missing value imputation, dimensionality reduction using Principal Component Analysis (PCA), exploratory data analysis, model training, and performance evaluation are all included in the extensive forecasting pipeline that was created. To ensure a fair evaluation of their predictive power, six supervised regression algorithms—Linear Regression, Ridge Regression, Lasso Regression, K-Neighbors Regression, Decision Tree Regression, and Gradient Boosting Regression—were created and contrasted using comparable datasets. Widely used regression metrics, such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R2 Score), were used to assess each model's performance. The K-Neighbors Regression model regularly outperforms the other regression models and achieves the highest prediction accuracy by generating smaller prediction errors and a higher R2 Score, according to experimental results. The comparative study shows that the reliability of wind power forecasting is greatly increased when a suitable machine learning technique is used in conjunction with efficient data preparation and feature optimization. In addition to supporting intelligent energy management, optimal scheduling of renewable energy resources, smart grid operation, and sustainable power generation, the proposed framework provides a feasible, scalable, and computationally efficient way for forecasting active wind power. In order to further increase prediction accuracy and operational efficiency in renewable energy applications, the developed system also offers a solid basis for future improvements involving bigger datasets, sophisticated ensemble learning techniques, hybrid machine learning architectures, real-time forecasting systems, and clever optimization strategies.

References

1. "Forecasting of Wind Turbine Output Power Using Machine Learning," H. Rashid, W. Haider, and C. Batunlu, Proceedings 10.1109/ACIT49673.2020.9208852, 10th International Conference on Advanced Computer Information Technologies (ACIT), Deggendorf, Germany, 2020, pp. 396–399.
2. "A Novel K-Nearest Neighbor-Based Machine Learning Approach for Wind Power Prediction Using Meteorological Data," Energy Conversion and Management, vol. 135, pp. 434–444, 2017; M. Yesilbudak, S. Sagiroglu, and I. Colak.
3. "Short-Term Wind Power Forecasting Using Genetic Programming Based Ensemble Neural Networks," Energy Conversion and Management, vol. 134, pp. 361–372, 2017; A. Zameer, J. Arshad, A. Khan, and M. A. Z. Raja.
4. "Machine Learning-Based Prediction of Wind Power Losses Caused by Icing Conditions," S. Scher and J. Molinder, IEEE Access, vol. 7, pp. 129421–129429, 2019.
5. "Artificial Intelligence-Based Analysis and Prediction of Floating Wind Turbine Performance," Ocean Engineering, vol. 175, pp. 207–216, 2019; X. Jiang, S. Day, D. Clelland, and X. Liang.
6. Z. Guo, W. Zhao, H. Lu, and J. Wang, "Multi-Step Wind Speed Forecasting Using Empirical Mode Decomposition and Neural Networks," Renewable Energy, vol. 50, pp. 241–249, 2012.
7. S. Baran, "Bayesian Model Averaging for Probabilistic Wind Speed Forecasting," Computational Statistics & Data Analysis, vol. 75, pp. 227–238, 2014.
8. "Data Mining Techniques for Wind Power Forecasting," L. Fugon, J. Juban, G. Kariniotakis, and B. Delille, Proceedings of the European Wind Energy Conference (EWEC), Brussels, Belgium, 2008.
9. Y. Liu, Z. Zhang, and X. Wang, "Machine Learning Techniques for Wind Power Forecasting: A Comprehensive Review," Renewable and Sustainable Energy Reviews, vol. 170, Art. no. 112967, 2022.
10. H. Wang, L. Chen, and Y. Li, "Comparative Study of Regression Algorithms for Renewable Energy Forecasting," Energy Reports, vol. 9, pp. 1485–1499, 2023.
11. "Artificial Intelligence Techniques for Smart Energy Systems: A Review," Journal of Cleaner Production, vol. 430, Art. no. 139708, 2024, M. Ferrag, L. Maglaras, H. Janicke, and A. Ahmim.
12. R. Sharma, A. Kumar, and S. Verma, "Wind Energy Prediction Using Ensemble Machine Learning Models," Engineering Applications of Artificial Intelligence, vol. 127, Art. no. 107366, 2024.
13. "Hybrid Machine Learning Framework for Wind Power Prediction in Smart Grids," J. Singh, P. Patel, and M. Desai, Applied Energy, vol. 367, Art. no. 123546, 2024.

14. "Principal Component Analysis-Based Feature Reduction for Renewable Energy Forecasting," K. Zhang, Y. Sun, and X. Liu, Sustainable Energy Technologies and Assessments, vol. 66, Art. no. 103885, 2024.
15. "Machine Learning Regression Models for Accurate Wind Power Estimation," S. Gupta and V. Sharma, IEEE Access, vol. 12, pp. 103421–103438, 2024.
16. "Performance Evaluation of Regression Algorithms for Renewable Energy Prediction," Energy AI, vol. 17, Art. no. 100412, 2025, D. Kumar, R. Patel, and S. Mishra.
17. Y. Chen, H. Zhao, and L. Wu, "Artificial Intelligence-Based Wind Energy Forecasting for Smart Grid Applications," Applied Soft Computing, vol. 154, Art. no. 111421, 2025.
18. "Advanced Machine Learning Models for Wind Power Generation Forecasting," P. Roy, S. Banerjee, and A. Das, Expert Systems with Applications, vol. 259, Art. no. 125109, 2025.
19. "Comparative Analysis of Machine Learning Algorithms for Renewable Energy Prediction," M. Ahmed, S. Ali, and F. Rahman, Journal of Computer and Information Sciences, King Saud University, vol. 37, no. 2, pp. 521–536, 2025.
20. "Data-Driven Machine Learning Techniques for Intelligent Wind Power Forecasting," T. Verma, N. Singh, and A. Joshi, Energy Reports, vol. 13, pp. 2250–2267, 2025.