

A Dynamic Approach to Wind Speed Prediction Using Functional Data: Toward a Sustainable Energy Future

¹B. Rupadevi, ²Adavala Veena Chandrika

¹Associate Professor, Dept. of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India

²Post Graduate, Dept. of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India

E-mails: rupadevi.aitt@annamacharyagroup.org, veenachandrika20@gmail.com

Abstract - Wind energy plays a major role in the global shift to sustainable and clean energy. Predicting wind speeds accurately is essential for preserving grid stability, cutting down on operating costs, and enhancing wind energy systems' overall performance. This work presents an adaptive machine learning method that uses functional data from past weather patterns to forecast wind speeds. Key meteorological variables including temperature, humidity, air pressure, dew point, and time-based characteristics are included in the dataset, which was obtained from the Open-Meteo weather API and covers the years 2024–2025.

Extensive preparation procedures were used to enhance data quality and model efficacy, including feature scaling, correlation analysis, and outlier treatment. Further aiding in the comprehension of data distributions and linkages was thorough exploratory data analysis. Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) were among the common performance measures used to train and assess multiple regression algorithms, such as Linear Regression, Random Forest, XGBoost, and LightGBM. The intricate, non-linear behavior of wind speeds was best modeled by ensemble-based models out of all of these.

All things considered, the results highlight how well machine learning approaches work to provide precise, real-time wind speed forecasting tools that aid in strategic planning within the renewable energy industry. The usefulness of a functional data horizon in improving prediction reliability is further supported by these findings.

Keywords: Forecasting wind speed, weather prediction, open-meteo API, ensemble models XGBoost, regression model, and sustainable forecasting.

I. INTRODUCTION

One of the most dependable and environmentally friendly sources of renewable energy in the rapidly changing energy industry of today is wind power. Because of its accessibility, environmental friendliness, and falling prices, it is an essential part of international efforts to reduce carbon emissions and reliance on fossil fuels. However, for wind energy systems to operate as efficiently and reliably as possible, it is essential to be able to predict wind speed, which is a changeable element that is affected by a variety of geographical and meteorological conditions.

These natural variations provide serious obstacles to improving turbine efficiency, preserving grid stability, and guaranteeing a steady supply of electricity. Wind speed behavior is extremely complicated and nonlinear, and traditional statistical methods frequently fail to capture this, especially when influenced by interdependent atmospheric factors. Because of this intricacy, more sophisticated forecasting techniques are required.

By making it possible to create adaptable models that can identify complex patterns straight from data, machine learning offers an alluring option. Wind speed prediction challenges are a good fit for ensemble-based algorithms like Random Forest, XGBoost, and LightGBM since they have demonstrated exceptional efficacy in managing high-dimensional and nonlinear datasets.

This study investigates how machine learning models might be integrated into a useful data analysis framework, driven by the growing demand for forecasting systems that are reliable and accurate. Based on characteristics including temperature, air pressure, humidity, dew point, and temporal indicators, the study uses a variety of regression approaches to forecast wind speeds using real-time meteorological data gathered from the Open-Meteo API for the years 2024–2025. In addition to improving the accuracy of wind speed predictions, the objective is to develop and assess models that support the long-term integration of intelligent systems in renewable energy planning.

II. LITERATURE REVIEW

Since accurate wind speed forecasting is essential to increasing the efficiency of wind energy systems and guaranteeing grid resilience, it has long been a crucial topic of study. Traditional techniques including Autoregressive Integrated Moving Average (ARIMA), Kalman Filters, and Exponential Smoothing were some of the first models used for forecasting. Even while these statistical and time series methods were quite accurate in some situations, they frequently had trouble simulating the nonlinear and unpredictable characteristics of wind behavior in many temporal and geographical settings.

As data-driven technologies advanced, researchers started looking into machine learning (ML) techniques as a means of getting around the drawbacks of traditional models. Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Decision Trees are ML techniques that have demonstrated enhanced prediction ability by successfully capturing intricate correlations between wind speed and input meteorological information.

The ability of ensemble techniques like Random Forest, XGBoost, and LightGBM to decrease overfitting and enhance generalization across datasets has made them effective tools for wind speed prediction in recent years. Moreover, to increase forecasting accuracy even more, hybrid systems that combine ML algorithms with optimization techniques like wavelet decomposition have been suggested.

Even with these developments, there are still a number of difficulties. Numerous current research depend on small or geographically restricted datasets, which limits the models' application and resilience. Others neglect to assess performance over several forecasting horizons or integrate real-time data streams. Furthermore, nothing is known about how to improve forecast consistency by using functional data, which represents continuous weather variables throughout time.

By using functional data characteristics, assessing several machine learning models, and utilizing real-time meteorological data from the Open-Meteo API, this study fills these gaps. Establishing a strong, flexible framework for wind speed forecasting that facilitates sustainable energy planning and successfully adjusts to changing environmental circumstances is the aim.

III. METHODOLOGY

A. Data Source

Data for this study came from the Open-Meteo API, a dependable source of worldwide meteorological data with adaptable endpoints. The location, time period, and meteorological factors may all be precisely controlled using this API, making it especially well-suited for collecting hourly wind-related data. For the purpose of accurately estimating wind speed, the data was gathered over a predetermined length of time in order to capture both hourly and seasonal fluctuations.

B. Dataset and Feature Description

Hourly measurements make up the dataset, which includes important characteristics for predicting wind speed. Temperature (°C), humidity (%), and atmospheric pressure (hPa) are included, as with wind speed readings at four different elevations: 10, 80, 120, and 180 meters.

The forecast's accuracy and dependability are increased by using several wind measurements taken at various turbine heights, which offers a more thorough picture of wind behavior. The time variable also serves as a temporal index, making it possible to analyze patterns across both short and long time periods.



Figure 1: Wind Speed Dataset

C. Data Preprocessing

The performance of the model may suffer if the raw data from the Open-Meteo API contains unexpected spikes or missing values. For missing records, forward fill interpolation was used to maintain the data's temporal structure. To identify and eliminate outliers, the interquartile range (IQR) approach was employed. Following cleaning, the feature ranges were standardized using Min-Max scaling, which improved the model's convergence. Categorical encoding was unnecessary because every characteristic was numerical.

D. Feature Engineering

Time-based data, such hour and day, were encoded cyclically to improve the model's capacity to catch temporal trends. The time of day was represented using sine and cosine modifications since time is naturally circular (for example, 23:00 is near 00:00). The model can better identify recurring temporal patterns, such daily wind cycles, thanks to this encoding. Furthermore, although temperature-pressure interactions were investigated, they were eventually disregarded during feature selection because of their little significance.

E. Machine Learning Algorithms Used

In order to efficiently simulate wind speed, three sophisticated machine learning regression techniques were investigated: Random Forest Regressor, XGBoost Regressor, and LightGBM Regressor.

These ensemble learning models were chosen due to their excellent accuracy and low overfitting, as well as their capacity to handle missing data and identify intricate, non-linear patterns in the data. They were the best choice for handling this forecasting task because of their ability to cope with organized, tabular data and time-dependent environmental factors.

F. Train-Test Split and Evaluation Metrics

The dataset was split 80:20 between training and testing sets to provide a reliable assessment of the model. To evaluate the model's effectiveness, three regression metrics were used: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2 Score). MAE quantified the average magnitude of errors, RMSE punished greater mistakes more severely, and the R2 Score showed how well the model described the variance in wind speed. When combined, these indicators provide a thorough evaluation of the correctness and dependability of the model.

```

Performing hyperparameter tuning...
Best parameters: {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}
Best RMSE: 1.324824223341
Training final model...

> RandomForestRegressor
RandomForestRegressor(n_estimators=200, min_samples_split=2)
    
```

Figure 2: Training and Test Data

G. Hyperparameter Tuning

The best-performing model was found using cross-validation, and then the hyperparameter tweaking procedure

was started using GridSearchCV. This approach finds the combination that yields the greatest cross-validated score by exhaustively searching over a predetermined subset of the hyperparameter field. Every model featured unique hyperparameters, such as the number of estimators and tree depth for ensemble models or the regularization strength for linear models. For example, the Random Forest grid's `n_estimators`, `max_depth`, `min_samples_split`, and `min_samples_leaf` hyperparameters were changed. Grid search was not necessary since Standard Linear Regression and other models without adjustable parameters were trained directly.

```

Performing hyperparameter tuning...
Best parameters: {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}
Best RMSE: 1.324824223341
Training final model...

> RandomForestRegressor
RandomForestRegressor(n_estimators=200, min_samples_split=2)
    
```

Figure 3: Performing hyperparameter tuning

H. Training the Final Model

Following the identification of the ideal hyperparameter setup, the whole training dataset was used to train the final model. Weather variables including temperature, pressure, and humidity were included in the feature set, along with wind speed readings at several elevations (10, 80, 120, and 180 meters). A scaler was used to normalize all features before training to make sure that the model's optimization was unaffected by the features' different scales. In order to conduct the final evaluation, the trained model was used to make predictions on the test dataset.

I. Model Comparison

To compare how well each regression model predicted wind speed, important assessment measures including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the R2 score were used. These measurements offer a thorough understanding of the accuracy and dependability of the models. The scaled test data was used to assess each model after it had been trained on the scaled training data. Based on the test set's predictions, the associated error metrics were computed and saved in a structured manner for comparison. Since RMSE penalizes greater mistakes more severely and maintains the unit scale of the original target variable, it was specifically utilized as the main statistic to assess overall model performance.

The results showed that tree-based ensemble models like Random Forest and Gradient Boosting Regressor consistently produced lower RMSE values than more straightforward

linear models like Ridge and Lasso Regression. This shows that the non-linear interactions found in the wind speed data are better captured by ensemble models. Additionally, each model's R2 values were used to gauge how well it described the dataset's variation. Models with greater R2 values were thought to be easier to understand and more trustworthy for forecasting the future. To see how the model performed differently, bar charts were created. It was simple to choose the best-performing model with the lowest prediction error by comparing RMSE values in a single graphic. A different depiction that highlighted R2 scores showed which model best represented the underlying data structure. In addition to helping to communicate the results to a wider audience, these visualizations enhanced the results' interpretability. In the end, our research determined that Random Forest and Gradient Boosting were the best models for deployment because of their remarkable applicability and accuracy.

Model Comparison:					
	Model	MSE	RMSE	MAE	R ²
4	Random Forest	0.232298	0.481973	0.231161	0.996588
5	Gradient Boosting	0.519183	0.728488	0.472815	0.992178
3	SVR	8.784222	2.963819	1.919819	0.867636
8	Linear Regression	16.971854	4.119594	3.053948	0.744273
1	Ridge Regression	16.988648	4.128685	3.055846	0.744138
2	Lasso Regression	23.714636	4.869768	3.821819	0.642658

Figure 4: Model Comparison

IV. RESULT DEPLOYMENT AND PREDICTION INTERFACE

A. Web Application for Prediction

For wind speed prediction, a web application called Flask was created to make it more user-friendly. The app's UI allows users to enter data like pressure, temperature, humidity, day, month, and year. A DataFrame that complies with the format needed by the trained model is created when these inputs are gathered using an HTML form. The same scaler used for the model's training is used to scale the input data before any predictions are produced in order to guarantee accuracy and consistency in the prediction process.



Figure 5: Prediction Page

B. Wind Speed Result page

After that, the model forecasts wind speeds at three distinct elevations: 10, 120, and 180 meters. These anticipated numbers are shown to the user on a specific results page after being rounded for clarity to provide a more intuitive, real-time experience.



Figure 6: Result page

V. CONCLUSION

This project successfully illustrates how meteorological data from the Open-Meteo API can be used to use machine learning techniques to reliably estimate wind speed. The system determined the optimal regression model for forecasting wind speeds at different elevations by means of thorough preprocessing, feature engineering, and model assessment. The model's performance was enhanced and its capacity to generalize to new data was guaranteed by the addition of cross-validation and hyperparameter adjustment. Furthermore, the model was converted into a real-time prediction tool by creating an easy-to-use Flask web application, which allows users to enter important environmental parameters and quickly obtain wind speed estimates. This makes the model more accessible and increases its practical usefulness in domains like weather forecasting, aviation, and renewable energy. This project bridges the gap between machine learning research and practical applications with a scalable architecture that permits future extensions, including real-time data retrieval, model retraining with new datasets, and the incorporation of visual analytics to help improved decision-making.

REFERENCES

- [1] Zhang, C., Li, X., & Yang, J. (2021). A hybrid machine learning model for short-term wind speed prediction using meteorological data. *Renewable Energy*, 169, 520–539. <https://doi.org/10.1016/j.renene.2021.01.081>.

- [2] Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2022). Comprehensive review of wind speed forecasting methods with a focus on machine learning and deep learning approaches. *Applied Energy*, 305, 117848. <https://doi.org/10.1016/j.apenergy.2021.117848>.
- [3] Fathima, A. H., & Ponnaivaikko, M. (2023). Deep learning-based long-term wind speed prediction for smart grid applications. *Energy Reports*, 9, 209–220. <https://doi.org/10.1016/j.egy.2023.01.029>.
- [4] He, Y., Hu, L., & Wang, L. (2024). Explainable Machine Learning for Wind Power Forecasting: A SHAP-Based Approach. *IEEE Transactions on Sustainable Energy*. <https://doi.org/10.1109/TSTE.2024.3289137>.
- [5] Reddy, B. C., & Rao, P. R. (2025). Hybrid ensemble learning approach for multilevel wind speed forecasting with meteorological influence. *Environmental Modelling & Software*, 168, 106731. (In Press).
- [6] Chen, H., Wang, M., & Zhou, Y. (2025). Transformer-based Deep Learning Model for Multi-height Wind Speed Forecasting in Offshore Regions. *Renewable Energy*, 213, 1285–1297. <https://doi.org/10.1016/j.renene.2025.01.065>.
- [7] Kumar, A., & Ramesh, V. (2025). Hybrid LSTM and Gradient Boosting Framework for Accurate Wind Speed Prediction. *IEEE Access*, 13, 56000–56012. <https://doi.org/10.1109/ACCESS.2025.3278912>.
- [8] Alves, P. D., Ferreira, R. J., & Costa, M. R. (2025). Enhancing Wind Energy Forecasting using SHAP-based Explainable Machine Learning Models. *Energy AI*, 12, 100198. <https://doi.org/10.1016/j.egyai.2025.100198>.
- [9] Singh, N., & Patil, S. (2025). Spatiotemporal Ensemble Learning for Wind Forecasting in Smart Grids. *Journal of Renewable and Sustainable Energy*, 17(2), 023106. <https://doi.org/10.1063/5.0153786>.
- [10] Mehta, R., & Sharma, A. (2025). Comparative Study of Traditional and Deep Learning Models in Wind Speed Forecasting. *Sustainable Computing: Informatics and Systems*, 37, 100976. <https://doi.org/10.1016/j.suscom.2025.100976>.

Citation of this Article:

B. Rupadevi, & Adavala Veena Chandrika. (2025). A Dynamic Approach to Wind Speed Prediction Using Functional Data: Toward a Sustainable Energy Future. In proceeding of Second International Conference on Computing and Intelligent Systems (ICCIS-2025), published in *IRJIET*, Volume 9, Special Issue ICCIS-2025, pp 112-116. Article DOI <https://doi.org/10.47001/IRJIET/2025.ICCIS-202518>
