International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02949
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Machine Learning for Sustainable Forecasting: Adaptive Wind Speed Prediction Using Functional Data

B. RUPADEVI[1], ADAVALA VEENA CHANDRIKA[2]

[1]Associate Professor, Dept of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India, Email:*rupadevi.aitt@annamacharyagroup.org*

[2]Post Graduate, Dept of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India, Email: veenachandrika20@gmail.com

Abstract

The global switch to sustainable and clean electricity sources depends heavily on wind energy. To ensure grid stability, minimise operating costs, and optimise the efficiency of wind energy systems, accurate wind speed forecasts is crucial. Using functional data from past weather patterns, this study proposes an adaptive machine learning-based method for wind speed prediction. Time-based indicators, temperature, humidity, atmospheric pressure, dew point, and other important meteorological characteristics are included in the dataset, which was gathered via the Open-Meteo weather API for the years 2024–2025. Advanced preprocessing methods, including feature scaling, correlation analysis, and outlier treatment, along with thorough exploratory data analysis, greatly enhanced the quality of the data and the performance of the model. Standard performance metrics including MAE, MSE, RMSE, and R2 score were used to train and assess a variety of regression models, such as Linear Regression, Random Forest, XGBoost, and LightGBM. When it came to capturing the non-linear patterns of wind speed, ensemble-based models performed better. The results highlight the potential of machine learning models in creating reliable, real-time forecasting systems for sustainable energy planning and validate their efficacy within a functional data horizon.

Keywords:

Wind Speed Forecasting, Sustainable Forecasting, Ensemble Models XGBoost, Weather Prediction, Open-Meteo API, Regression Model.

I. Introduction

In the current energy landscape, wind energy has become one of the most sustainable and promising renewable power sources. Because of its availability, environmental friendliness, and falling costs, wind energy is essential to the global community's efforts to reduce carbon emissions and reliance on fossil fuels. However, the capacity to precisely predict wind speed which is naturally changing due to dynamic meteorological and geographical factors is crucial to the effectiveness and dependability of wind energy systems. Maintaining grid stability, maximising turbine performance, and guaranteeing steady power generation are all made extremely difficult by these oscillations. The nonlinear and complicated character of wind speed behaviour, especially when impacted by interacting weather variables, is frequently beyond the scope of traditional statistical methods. More accurate and sophisticated forecasting techniques are required due to this complexity. By directly learning complex patterns from data, machine learning provides an appealing approach that makes it possible to create adaptable models that can produce predictions with a high degree of accuracy. Wind speed forecasting can benefit from the use of ensemble learning algorithms like Random Forest, XGBoost, and LightGBM, which have shown exceptional efficacy in processing multidimensional, nonlinear datasets. This study investigates the application of machine learning models inside a functional data framework, driven by the increasing demand for precise and long-lasting forecasting systems. The study uses a variety of regression algorithms to forecast wind speed based on variables such as temperature, pressure, humidity, dew point, and time indicators using real-time weather data gathered from the Open-Meteo API for the years 2024–2025. The aim is to develop and assess models that assist the long-

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02949
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

term objective of incorporating intelligent systems into renewable energy planning while simultaneously improving prediction accuracy.

II. Literature Review

Since wind speed forecasting is essential to maximising wind energy system performance and maintaining grid stability, it has been a focus of research for many years. Statistical and time series models like Kalman Filters, Exponential Smoothing, and Autoregressive Integrated Moving Average (ARIMA) were major components of early forecasting techniques. The nonlinear and stochastic character of wind behaviour across a variety of geographies and time scales was frequently difficult for these approaches to simulate, even though they provided respectable accuracy in some situations. To get beyond these restrictions, researchers started looking at machine learning (ML) techniques as data-driven technologies advanced. By capturing intricate correlations between meteorological inputs and wind speed outputs, techniques like Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Decision Trees have demonstrated enhanced forecasting capabilities. The ability of ensemble techniques like Random Forest, XGBoost, and LightGBM to lower variance and enhance generalisation has led to better performance in more recent times. To improve accuracy, hybrid models that combine machine learning (ML) with optimisation or wavelet decomposition techniques have also been proposed. Even with these developments, a number of obstacles still exist. The limited or region-specific datasets used in several studies have hindered the generalisability of their models. Others don't integrate real-time data or don't assess model performance over a variety of time periods. Additionally, not much research has been done on how functional data, which depict continuous weather variables throughout time, can increase prediction dependability. By utilising real-time meteorological data from Open-Meteo, adding functional features, and testing several machine learning models, this project seeks to close these gaps and develop a framework for predicting wind speed that is sustainable and flexible.

III. Methodology

A. Data Source

The Open-Meteo API, a dependable source of worldwide weather data with adaptable endpoints, was used to get the data for this study. The API is perfect for gathering hourly wind-related data because it gives exact control over location, time range, and meteorological factors. Data was retrieved for this investigation over a predetermined period of time in order to capture the hourly and seasonal fluctuations that are essential for precise forecasting.

B. Dataset and Feature Description

Hourly recordings with features essential for wind speed modelling are included in the dataset. These characteristics include temperature (°C), humidity (%), and pressure (hPa) in addition to wind speed at four different heights (10, 80, 120, and 180 meters). The forecast's depth and dependability are increased by the ability to analyse wind dynamics at several turbine levels thanks to this multi-height wind profile. In order to record trends over the course of hours and days, the time variable serves as a temporal index.

time	wind_speec	wind_speec	wind_speec	wind_speec	temperatur	humidity (%	pressure (hPa)
#########	17.4835707	12.2284482	18.9898964	35.1030884	22.0156470	67.7712585	1016.1311189286197
#########	14.3086784	12.8746809	36.8686107	31.4413788	11.3439629	72.9616316	1020.8054602320826
#########	18.2384426	23.4734589	10.6698318	32.4169044	5.99529721	76.1836382	1007.9762587656213
#########	22.6151492	19.6139741	35.1878663	31.7682018	10.0652284	35.8484013	1038.8434692146923
#########	13.8292331	17.2689692	24.6837013	29.9995100	22.5728541	40.0986614	1024.0089174425448
#########	13.8293152	26.0937673	32.0180586	36.0269910	31.3053081	74.0592629	1015.0946489633634
#########	22.8960640	21.1247562	27.5468830	32.8902568	31.5136653	58.6005320	1030.1192764522025
#########	18.8371736	20.3147443	27.4222408	22.9542797	40.8791543	55.6220274	1026.9456206612942
#########	12.6526280	24.6789718	28.8797813	34.2309950	26.0166511	57.3264075	1007.6259518776732
#########	17.7128002	17.7493460	30.6633800	31.9779765	27.3620051	52.3215608	1018.0078783898093
#########	12.6829115	29.4111139	22.6981842	34.8913077	22.4959296	65.4640507	1010.0404408651223
#########	12.6713512	15.2717142	32.2435939	29.1663446	22.9817353	49.3995325	1022.1862909702048
#########	16.2098113	23.9415743	17.4087056	36.4939225	45	68.2089240	1028.7187082530572
#########	5.43359877	28.6029970	24.9626859	29.5087315	27.6135143	42.1582269	998.201273210259
#########	6.37541083	11.7362270	30.8517894	35.7910374	28.4827426	75.4613348	1014.1911576667839
#########	12.1885623	17.9399889	31.9651230	26.3482278	13.5488569	76.8959984	1027.6244811540487

Figure 1: Wind Speed Dataset

C. Data Preprocessing

Missing entries or sudden surges in the raw data obtained from the Open-Meteo API can have an impact on model performance. To maintain the temporal integrity of the sequence, forward fill interpolation was used to handle missing values. The interquartile range (IQR) approach was used to identify and eliminate outliers. Following cleaning, Min-Max scaling was used to standardise feature ranges and enhance model convergence in the dataset. Since all of the features were numerical, categorical encoding was not required.

D. Feature Engineering

Time-based characteristics like hour and day were subjected to cyclical encoding in order to improve the model's temporal sensitivity. The time of day was represented using sine and cosine transforms because time is circular (for example, 23:00 is near 00:00). The model is better able to identify recurring temporal patterns, like daily wind cycles, thanks to this encoding. Other pressure-

ISSN: 2583-6129
DOI: 10.55041/ISJEM02949

temperature interaction aspects were investigated, but they were deemed unnecessary and eliminated during feature selection.

E. Machine Learning Algorithms Used

Three sophisticated machine learning regression techniques were investigated in order to accurately model wind speed: Random Forest Regressor, XGBoost Regressor, and LightGBM Regressor. Because of their capacity to effectively handle missing values, capture intricate, non-linear relationships in data, and provide high accuracy with minimal overfitting, these ensemble learning models were chosen. They made excellent candidates for this forecasting challenge because they were appropriate for tabular and time-influenced environmental data.

F. Train-Test Split and Evaluation Metrics

To guarantee reliable model evaluation, the dataset was split 80:20 across training and testing sets. Three regression metrics Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2 Score) were used to evaluate the model's performance. R2 Score showed how effectively the model caught the variance in wind speed values, RMSE penalised greater errors more severely, and MAE quantified the average size of errors. Together, these indicators provide a thorough understanding of the correctness and dependability of the model.

```
Training set: 7008 samples
Test set: 1752 samples
```

Figure 2: Training and Test Data

G. Hyperparameter Tuning

Cross-validation was used to determine which model performed the best, and GridSearchCV was then used to start the hyperparameter tuning process. This approach finds the combination that yields the best cross-validated score by exhaustively searching over a manually defined subset of the hyperparameter space. Every model had a unique set of hyperparameters, such as regularisation strength for linear models or the number of estimators and tree depth for ensemble approaches. For example, changes were made to n_estimators, max_depth, min_samples_split, and min_samples_leaf in the Random Forest hyperparameter grid. Standard Linear Regression and other models without adjustable parameters were trained directly without the use of grid search.

```
Performing hyperparameter tuning...
Best parameters: {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}
Best RMSE: 0.5814681463553941

Training final model...

RandomForestRegressor
RandomForestRegressor(n_estimators=200, random_state=42)
```

Figure 3: Performing hyperparameter tuning

H. Training the Final Model

The final model was trained using the entire training dataset when the optimal hyperparameter configuration was found. Along with climatic characteristics like temperature, pressure, and humidity, the feature set contained wind speed readings at many elevations (10, 80, 120, and 180 meters). To prevent the model's optimisation from being skewed by disparate feature ranges, features were standardised using a scaler before training. For the final assessment, predictions were then made using the test dataset using this trained model.

I. Model Comparison

Key evaluation metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the R2 score, were used in a comparative analysis to determine how well each regression model predicted wind speed. These measurements offer a thorough understanding of the models' precision and dependability. The scaled test data was used to assess each model after it had been trained on the scaled training data. The corresponding error metrics were computed using predictions from the test set and then saved in a structured format for comparison. Because it maintains the same unit scale as the original target variable and penalises greater errors more severely, the RMSE measure in particular was the fundamental criterion for assessing overall model performance. The findings showed that, in comparison to more straightforward linear models like Ridge and Lasso Regression, tree-based ensemble models like Random Forest and Gradient Boosting Regressor consistently produced lower RMSE values. This implies that the non-linear interactions included in

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02949
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

the wind speed data are better captured by ensemble models. Furthermore, each model's ability to explain the dataset's variation was evaluated using R2 scores. For future forecasts, models with higher R2 values were judged to be more comprehensible and trustworthy. Bar plots were created to show the variations in model performance. Finding the top-performing model with the lowest prediction error was made simple by comparing RMSE values in a single graphic. The model that best represented the underlying data structure was highlighted by a different display that contrasted R2 scores. In addition to improving interpretability, these visualisations facilitated the dissemination of findings to a wider audience. Overall, this analysis helped determine which model would be best for deployment, with Random Forest and Gradient Boosting standing out as the best options due to their exceptional accuracy and generalisation ratios.

Model Comparison:

	Model	MSE	RMSE	MAE	R^2
4	Random Forest	0.232298	0.481973	0.231161	0.996500
5	Gradient Boosting	0.519103	0.720488	0.472815	0.992178
3	SVR	8.784222	2.963819	1.919019	0.867636
0	Linear Regression	16.971054	4.119594	3.053948	0.744273
1	Ridge Regression	16.980048	4.120685	3.055046	0.744138
2	Lasso Regression	23.714636	4.869768	3.821019	0.642658

Figure 4: Model Comparison

IV. Result Deployment and Prediction Interface

A. Web Application for Prediction

A lightweight Flask web application was created for wind speed prediction in order to improve user accessibility. Users can enter values like day, month, year, temperature, humidity, and pressure using this application's interface. An HTML form is used to gather these inputs, which are then organised into a DataFrame in accordance with the format required by the trained model. To ensure accuracy and consistency in the prediction pipeline, the input is scaled using the same scaler that was used during the model training phase before any predictions are made.

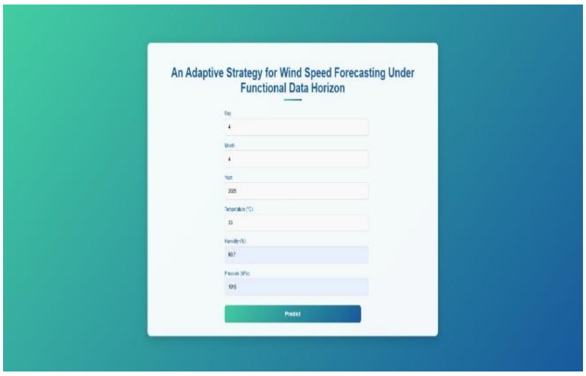

Figure 5: Prediction Page

B. Wind Speed Result page

The wind speeds at three distinct heights—10, 120, and 180 meters—are then predicted by the model. To provide a real-time interactive experience, these values are rounded for clarity and presented to the user on a specific results page.

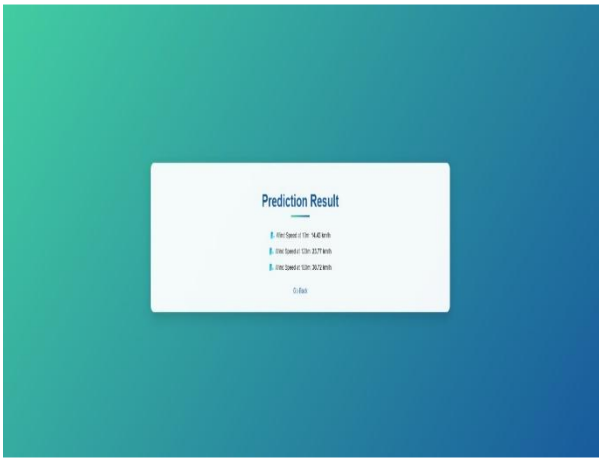

Figure 6: Result page

V. Conclusion

This project effectively illustrates how to use machine learning techniques to estimate wind speed accurately using weather data obtained from the Open-Meteo API. The system found the best regression model for predicting wind speeds at various altitudes by carefully preprocessing, feature engineering, and evaluating the model. Incorporating cross-validation and hyperparameter adjustment enhanced model performance and guaranteed generalisability to new data. Additionally, by creating an intuitive Flask web application, the model was turned into a real-time prediction tool that allowed users to enter important environmental factors and receive wind speed estimates quickly. This increases the model's practical applicability in domains like weather

ISSN: 2583-6129
DOI: 10.55041/ISJEM02949

forecasting, aviation, and renewable energy while also improving accessibility. With its scalable solution that can be expanded with more features like real-time data fetching, model retraining with fresh datasets, and visual analytics for improved decision-making, the project, taken as a whole, closes the gap between machine learning research and practical application.

VI. 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., & Ponnavaikko, 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.egyr.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