Explainable Stacked Ensemble Learning for Accurate Wind Power Forecasting Using SCADA Time-Series Data

MSc Research Project
Programme Name

SANKET SAWANT
Student ID: x23268077

School of Computing
National College of Ireland

Supervisor: Prof. Vikas Tomer

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Sanket Balasaheb Sawant
Student ID:	x23268077
Programme:	MSc Data Analytics
Year:	2025
Module:	MSc Research Project
Supervisor:	Prof. Vikas Tomer
Submission Due Date:	15th September 2025
Project Title:	Explainable Stacked Ensemble Learning for accurate wind power forecasting using Scada time series data
Word Count:	7287
Page Count:	20 Pages

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Sanket Sawant

Date: 15th September 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Explainable Stacked Ensemble Learning for Accurate Wind Power Forecasting Using SCADA Time-Series Data

Sanket Sawant
x23268077

Abstract

Wind power forecasting is the process of predicting electricity produced by wind turbines based on environmental and operating parameters which are usually gathered by SCADA (Supervisory Control and Data Acquisition) systems. The common issues of non-linear dependence, local anomalies and the absence of interpretability are faced by traditional forecasting methods like the physical model that is based on a numerical weather prediction and stand-alone statistical models, such as ARIMA or linear regressors. The proposed work is an attempt at overcoming these shortcomings given the encouraging results of the study to propose a stacked ensemble learning model through which four effective regressors namely Support Vector Regressor (SVR), Random Forest, AdaBoost, and XGBoost and are combined using a meta-learner, which improves the accuracy and robustness of the proposed model. Time-series feature engineering approaches such as the rolling of windows are exploited in this model. To verify the efficiency of the model 5-fold cross-validation strategy was used and the R 2 score of this approach was astonishing 0.9292, mean squared error (MSE) was 0.0009, and root mean squared error (RMSE) was 0.0293 and this was the best model using any baseline model. Also, Local Interpretable Model-agnostic Explanations (LIME) was used to enable transparency and to identify important SCADA features impacting predictions. The results can support the claim that the stacked ensemble is scalable and reliable as a possible tool to make real-timely energy management in smart grid settings both interpretable and have high predictive accuracy.

Keywords: Wind Power Forecasting, SCADA Data, Stacked Ensemble Learning, Time-Series Prediction, Explainable AI (LIME)

1 Introduction:

1.1 Aim of the study

The main aim of the study is to come up with a powerful, precise and comprehensible machine learning algorithm to predict the power generation of a wind turbine based on SCADA (Supervisory Control and Data Acquisition) data. Considering the nature of variability and complexity in the production of wind energy, the proposed study is aimed at exploiting advantages of multimodels of machine learning by using stacked ensemble strategy to improve their predictive performance and strength in generalization. The methodology incorporates the most important models, Support Vector Regressor (SVR), Random Forest, AdaBoost, and XGBoost integrated with each other or rather constituting the final ensemble through use of a meta-learner. Data SCADA time-series are preprocessed and engineered and trained on model of rolling windows to capture the dependency in time. In addition, the research involves explainable AI (LIME) to explain model predictions and highlight important SCADA characteristics and bring transparency in decision-making. Thanks to the deployment of 5-fold cross-validation, the reliability and robustness of the model is proved carefully. Finally, it is hoped that the work will add value to the body of knowledge in renewable energy forecasting due to the inherent feasibility, accuracy, and explanability in the proposed solution that would be suitable to implement into the real-time energy management systems. Not only does the results show better performance than individual models, but it also gives a practical direction to be applied by the wind farm operators and the smart grid systems.

1.2 Research Questions

How does a stacked ensemble model combining multiple machine learning algorithms improve the accuracy and robustness of wind power forecasting using SCADA data, and which SCADA features most significantly influence the predictions as revealed by explainable AI techniques (LIME)?

1.3 Objectives of the Research

The objectives of research for this report are:

1. Build and compare a stack ensemble machine learning model to create a more accurate and robust model to predict power output of the wind turbine based on SCADA using a combination of the SVR, Random Forest, AdaBoost, and XGBoost algorithms.

2. To use time-series feature engineering strategies, including rolling to efficiently discern temporal relationships within the wind power data to train models optimally to make short-term forecasts.

3. Use explainable AI methods (LIME) to interpret ensemble model predictions and best SCADA features driving power generation, thus increasing transparency and model reliability to energy management systems in the real world.

1.4 Outline of the Report

- This report has been organized in the following way: Introduction: presents the research problem, purpose, goals as well as the importance of forecasting wind power based on machine learning and explainable AI.
- Literature Review: provides a review of the existing forecasting methods, ensemble learning techniques, and interpretability of energy prediction models.
- Research Methodology: Details of dataset, preprocessing, feature engineering and selection as well as implementation of explainable AI techniques.

- Design Specification: It lays out the system architecture and system workflow interface, starting with system data acquisition to ensemble prediction and interpretability.
- Implementation: Describes the code, the model training, the process of forecasting and visualization of the results obtained using LIME and time-series impending.
- Comparison: It involves comparison of the performances of the various models using quantitative measures and graphs showing how the stacking ensemble model outperformed the other models.
- Conclusion and Future Work: Summarizes key findings and proposes future directions like deep learning integration and real-time deployment.

2 Related Work:

2.1 Wind Power Forecasting: An Overview

The wind power forecasting refers to the practice of anticipating the level of electrical energy that can be produced by the wind turbines in the course of a given time period Kamani and Ardehali (2023). It is important in making power systems with renewable energy reliable, efficient and stable. Since wind energy is intermittent and variable in nature, proper forecasting is necessary so that the grid operators can match supply and demand, schedule reserves and minimise the costs of matching power imbalance or use of back-up fossil-fuel- based generation plants Chen et al. (2023). Such forecasts may be over several minutes to several days and are usually differentiated into very short-term (seconds or minutes), short- term (minutes to hours), medium-term (hours to days), and long-term (days to weeks) depending on the specific application Czapaj et al. (2022). Better forecasting helps us ensure that the operator of a wind farm sets its maintenance routine, optimises the operation of a turbine, becomes a more valuable player in energy markets, and can support real-time control Udo et al. (2024). The advantages are in economical returns, enhanced interconnection with the wind and enhanced decision plotting in trading energy Munoz et al. (2020). Those forecasting approaches that were still performed traditionally were physical or statistical. Physical models are also based on numerical weather predictions (NWP) and parameters depending on the turbine, however, calculate intensively and can fail to capture the local variability. Such statistical models as ARIMA and persistence models are simpler but do not have the capacity to model more complex non-linear dependency. Machine learning, as an advanced technology, has become a real option within the past few years; being able to learn patterns on historical data Sarker (2021) and adjusting itself to the dynamics of the wind behavior without having architected explicit physical models. ML models can deal with high amounts of data, handle miscellaneous inputs on the scale of environments and turbines, and become even more accurate as further data is acquired.

2.2 ML Techniques in Wind Power Prediction

Alkesaiberi et al. (2022) have suggested the data-driven method of enhancement of the accuracy of forecasting wind power production, this being of extreme importance considering that wind energy is inherently intermittent and variable. The research mainly emphasized on univariate wind power time-series prediction with good machine learning models such as Gaussian Process Regression (GPR), Support Vector Regression (SVR) of different kernels as well as ensemble learning approaches, such as Boosted Trees and Bagged Trees. The

Bayesian Optimization (BO) was used to optimize both sets of models to correctly allocate hyperparameters and improve accuracy. One of the factors that make the approach very strong is the incorporation of dynamic information which has been done through introduction of lagged values and other input features that include the wind speed and direction. This allowed the models to adequately come to the terms with temporal dependencies within the data. This methodology was tested with actual data in wind turbines in France, Turkey and Kaggle source. Findings showed that GPR and ensemble models with tunable parameters have better results compared with the other models particularly those improved by the use of lagged data and external input variables. The weakness of the study is, however, the fact that the study relies on past trends of data which can affect model flexibility in case of sudden changes in the environment or in situations where the turbine used has abnormalities that were not captured in the historical data.

According to Tarek et al. (2023), the authors intended to make the process of predicting wind power more descriptive and accurate by creating a number of regression ML models that were based on machines and deep learning, namely DNN, KNN, LSTM, Random Forest, Bagging, Gradient Boosting, and Averaging. The dataset contained four features including 50,530 instances after thorough data cleaning and data preprocessing procedures were used. Stochastic Fractal Search Outfitted with Particle Swarm Optimization (SFS- PSO) a new hybrid optimization algorithm was offered to overcome the traditional related issue of the LSTM network of suboptimal performance caused by poor hyperparameter configurations. Each model performance was compared based on five performance measures, i.e. MAE, NSE, MSE, R^2 and RMSE. The optimized LSTM with SFS-PSO is the best model when compared to the other models tested since it attained an R^2 of 99.99 % which means the model was almost perfect in its prediction ability. Although these are very encouraging findings, the method can be constrained by the complexity of computations and scalability of the algorithm and also the use of only one type of optimization process may not be generalizable to every forecasting setting.

Another study deals with the issue of forecasting wind power given by Karaman (2023), which is a complicated task but a multiscope predictive model is proposed that uses modern machine learning algorithms such as Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks. The authors added together two separate datasets: the first one included internal parameters, namely, wind speed, wind direction, and theoretical power, and active power, whereas the other one involved the external meteorological data to enhance the feature space. With the k-nearest neighbors (kNN) algorithm, missing data were imputed. Traditional statistical measures were chosen to evaluate the performance of the models as R^2, RMSE, MAE, and MSE. Of all the models, LSTM network has been shown to most accurately with R^2 of 0.9574, MAE of 0.0209, MSE of 0.0038, and RMSE of 0.0614, a sign that it is very successful in its forecasting skill. The fact that both internal and external aspects were included also enhanced predictive accuracy, however, certain limitations still exist, including the possibility to overfit on complicated model architectures and reliance on high quality and multi-source data that may not readily exist in all forecasting scenarios.

In He et al. (2022), the research is aimed at wind farm efficiency by trying to solve the problem of effectively predicting fatigue loads and power generation during yaw control, which is a major method in active wake control that is used to maximize the overall wind farm generation. In noting the possible vulnerability of the structure to the occurrence of yaw misalignment, the authors have suggested that a Support Vector Regression (SVR) based machine learning model can be used to predict and forecast both fatigue loads on a wind turbine associate critical components and the subsequent power mentioned. The methodology takes into consideration wake effects of upstream turbines and has a large variety of inflow conditions (and yaw conditions). This makes model training comprehensive and models robust. The SVR model was tuned to take the complicated interactions into consideration of operational conditions and turbine performance. The superiority of the suggested method was verified by the comparative outcomes comparing to another ML model, with the high regression coefficients and minimum deviation in the predictions made. It is interesting to note that the model performed better where the yaw angles were large and the wind speed high. Although simplification of the method holds future promises in advancing yaw optimization strategies in wind farms, one of the limitations is the fact that such method relies on accurate and large information on operations and its application may be limited in generating the models depending on the turbine type or site environment where they are used.

The review introduced by Dhiman and Deb (2020) covers all possible methods of wind speed forecasting with emphasis on machine learning methods and hybrid models designed to operate with time-series regression. Since the issue of temporal variability of the wind speed is an important one as far as applications like economic load dispatch, market clearing and grid scheduling are concerned, the authors of the research have made it clear that models in wind prediction be classified according to their forecasting horizon. This review tests supervised learning models (such as support vector regression, multiple linear regression, and neural networks) on global wind speed datasets and with the aim of determining their generalisation ability. Although such models perform well when it comes to the adaptation process of unseen data, issues such as overfitting present, particular in single models. To counter this, the review shows the increased applicability and level of effectiveness of ensemble and hybrid models which incorporates the combination of several forecasting techniques in order to make the method more accurate and reliable. The principal concluding points are that hybrid models are more reliable in making their predictions as compared to the solitary models. A particular limitation that has been described is the difficulty of ensemble systems to perform well in real-time forecasting because of the extra model complexity and computational overhead requirements.

In Hanifi et al. (2020), the authors endeavour to find a solution to the formidable challenge of uncertainty, and fluctuation of the wind, which adversely affects the successful large-scale incorporation of wind power into the grid, by undertaking a systematic review of the existing means of forecasting wind power sources. Their paper compares and contrasts how well they perform, how computationally demanding they are, their input characteristics and the range of time horizons over which they can be effectively applied to the application of physical models, statistical methods: time-series, and structural approaches, artificial neural networks and hybrid methods to a variety of stationary and nonstationary data sets. Several aspects that influence the accuracy of forecast constructs have also been discussed such as model complexity and metric of error measurement. The identified contribution is the establishment of a well-organized guide to help wind farm workers in choosing an appropriate forecasting method applicable to their operating interests and the cost of trade-off between accuracy and computer time. Also, the paper makes strategic suggestions on the future research line of action in the area. As far as the review provides a comprehensive and insightful picture of the field of forecasting, its limitation is based on the fact that it relies on the existing literature without providing empirical validation and comparative implementation, which could impede its potential to comprehensively observe the model performance in an evolving, real-life environment.

2.3 Stacked Ensemble Learning in Renewable Energy

To enhance the effectiveness of the short-term prediction of renewable energy production, mainly wind energy and photovoltaic, Huang et al., (2023) developed an adaptive stacking ensemble learning framework that can better the quality of forecasting. The research looks at the fundamental problem presented by the intermittence and variability of renewable energy sources that has a critical effect on the security and stability of power system operations. To curthilt these difficulties, the authors used a systematic process in which determination coefficient (R2) was used as the measure to assess the performance of the twelve candidate models and the best five performing models were chosen as base learners. Optimization of these models was further achieved using Bayesian hyper parameter tuning and diversification was done via the use of cross-validation. The last ensemble was constructed through a linear model where base models have been assigned weights according to lower cross-validation errors. When fitted on four seasonal datasets that included wind farms and solar stations, the suggested model performed better in terms of prediction accuracy and generalization than the baseline techniques. The possible weaknesses however include the model used only on the past data trends so it may not be as flexible when exposed in highly uncontrollable or untrapped environments and it will continuously have to be upgraded and retrained to constantly perform accordingly.

Khan et al. (2022) designed a Deep Stacked Ensemble model (DSE-XGB) to increase the precision or certainty of solar energy forecasting, an important operation to increase the number of renewables with the incorporation of the existing power grid. The research draws on the availability of growing volumes of high-resolution solar generation data, to create a data-driven methodology based on a combination of two deep learning techniques, namely, Artificial Neural Network (ANN) and Long Short-term Memory (LSTM) as baseline predicting models. The result is then a combination of these forecasts with the help of the Extreme Gradient Boosting (XGBoost) algorithm to final output, where challenges like variability on the basis of weather conditions and a pre-eminence in the predictive constancy of a single model are overcome. The model would be tested in four different solar datasets, and its interpretability would be enhanced by use of the SHAP (SHapley Additive exPlanations) framework, hence maximal clarity on feature contribution. Findings indicated that DSE-XGB model performed better than the individual ANN, LSTM, and Bagging models with an R2 improvement of 10-12 percent and also proved more consistent in the various weather conditions. Their possible limitation is, however, the fact that deep learning models coupled with XGBoost can be computationally expensive and time-consuming during training, which can further impact their use in the real-time application and performance in a resource-limited setting.

Alghamdi et al. (2023) suggested a generalized and unified framework towards doing accurate forecasting of both wind speed and solar radiation being some of the key sources of renewable energy facing the challenges of high computational complexity and insufficient generalisation by the traditional forecasting procedures. To optimize the parameters of a newly developed stacked ensemble learning model, the research proposed an innovative hybrid algorithm, which was named GABER (Al-Biruni Earth Radius (BER) + Genetic Algorithm (GA)), to parameterize this type of model. This non-linear optimization related to a hybrid was the key feature in ensuring that a high level of prediction accuracy was achieved and this was used to enhance the generalization capabilities of the model to varying forms of renewable energy sources. The model has been confirmed with the help of numerous experiments and statistical analysis, and it often proved its remarkable efficiency in prediction, stability, and generalization compared to current state-of-the-art forecasting and optimization methods. Nevertheless, the proposed method could have a shortcoming of the presence of enhanced algorithmic complexity that would be posed by the hybrid GABER optimizer, which would interfere with the scalability and performance of real-time or large- scale renewable power systems.

At last there is a study which is developed by Abdellaif et al. (2022) introduced a stacked ensemble learning algorithm, Stack-ETR, to solve photovoltaic (PV) power output instability problem due to high PV penetration in the modern power distribution networks by forecasting one-day-ahead output. The analysis was meant to increase the accuracy of prediction as a way of effective energy management and improved grid integration. The method combines three machine learning algorithms, Random Forest Regressor (RFR), Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost) as base learners, in which an Extra Trees Regressor (ETR) is the meta-learner to make a combination of their output. The model was tested against three actual PV systems (thin-film, monocrystalline and polycrystalline) by using dataset of four years meteorological data. It also showed the effectiveness of Stack-ETR over other ensemble models with reduction of RMSE by as much as 40.2 percent and reduction of MAE by as much as 47.2 percent relative to the standalone ETR model in all types of PV. Although the model exhibits excellent reliability in predicting and having flexibility in its adaptation, one of its weaknesses could be the fact that it will require quality, and long-term historical meteorological data that may not be readily available in any deployment location.

2.4 Explainable AI (XAI) in Wind Forecasting

Recent wind power predictive applications of XAI have received particular interest because of the trade-off between model predictive accuracy and model interpretation inherent in the current state of advanced AI forecasting. Although models of deep learning and machine learning (LSTM, GRU, ARIMA, Transformer, etc.) have proven to be able to increase the accuracy of wind power predictions dramatically, their so-called black-box characteristic leads to doubts in trustworthiness, transparency, and decision making in operations.

The authors Liao et al. (2024) were able to resolve this problem because they studied four model-agnostic XAI methods--Shapley Additive Explanations (SHAP), Permutation Feature Importance (PFI), Partial Dependence Plot (PDP), and Local Interpretable Model-Agnostic Explanations (LIME), which gives global and instance-level interpretation of the AI model as applied to wind power forecasting. They have suggested also the evaluation metrics to estimate the credibility of the interpretations. Their findings confirmed that the above methods would enable identification of important aspects affecting forecasts as well as determination of the levels of the reliability of interpretation by the users, making model selection easier.

Just the same, Yaseen et al. (2025) developed a new hybrid model, with the hybrid model being X-double LSTM that integrates LSTM with modified SHAP to improve the accuracy of the prediction as well as explaining how each input feature works in predicting time series data of wind power. The model showed better results than traditional models, i.e., singlelayer LSTM, GRU, ARMA, Transformer, and Bi-LSTM, on nine statistical measures of performance, viz., RMSE, MAE, R 2, and NSE, which were run on Google Colab utilizing TensorFlow and Keras. SHAP inclusion in this model allowed highlighting significant features influencing accuracy in the forecast, thus, enhancing its interpretability to not affect performance.

Kabir et al. (2024) have shown in an innovative usage that they combined XAI to determine the effect of rain on the aerodynamics of Horizontal-Axis Wind Turbines (HAWTs) in an application of Blade Element Momentum (BEM) theory. The modeling of S809 airfoil in several rain intensities and symbolic regression consequently displayed that rain influences each the lift and drag thus plunging power generation by 5.7 per cent to 7 per cent of the un-rained case at elevated liquid water content (LWC). To offer a lower-computation alternative, the necessity of conventional simulations was filled in by the use of the XAI framework based on machine learning models, which was used to predict the changes in aerodynamic coefficients with an R2 of 0.97.

In the last, Kim and Kim (2023) tackled the question of diversity of input into the models and aimed at its enrichment by, in particular, including such meteorological features of wind generation as the intensity of turbulence, the turbulent kinetic energy of the atmosphere, and atmospheric stability in ML models to achieve the goal of predicting the 10-minute average power and energy production during a day.

Table No: 1

Study	Proposed Method/Approach	Dataset Used	Results/Performance	Limitations
Alkesaiberi et al. (2022)	GPR, SVR (various kernels), Boosted Trees, Bagged Trees with Bayesian Optimization	Wind turbine datasets (France, Turkey, Kaggle)	GPR and ensemble models with lagged features achieved higher accuracy	Relies heavily on past trends; limited flexibility during sudden environmental changes
Tarek et al. (2023)	DNN, KNN, LSTM, RF, Bagging, Gradient Boosting; optimized LSTM using SFS-PSO	50,530 cleaned instances, 4 features	Optimized LSTM achieved $R^2 \approx 0.9999$, outperforming all models	High computational complexity; single optimization may not generalize
Karaman (2023)	ANN, RNN, CNN, LSTM; combined internal turbine + external meteorological data	Two datasets (internal turbine + external meteorological)	LSTM best: $R^2 = 0.9574$, RMSE = 0.0614	Risk of overfitting; requires multi-source high-quality data

He et al. (2022)	SVR model for fatigue load & power prediction during yaw control	Operational wind farm data (varied yaw & inflow conditions)	High regression accuracy; robust under large yaw angles and high wind speeds	Requires large, detailed operational data; may not generalize to all turbine types
Dhiman & Deb (2020)	Review of ML forecasting methods (SVR, MLR, NN, hybrid/ensemble)	Global wind speed datasets (literature-based)	Hybrid/ensemble models more reliable than single models	No empirical validation; real-time application limited by complexity
Hanifi et al. (2020)	Systematic review of physical, statistical, ANN, and hybrid methods	Literature on stationary & non-stationary datasets	Provided guidelines for selecting forecasting models; identified trade-offs	No experimental validation; conclusions rely on secondary sources
Huang et al. (2023)	Adaptive stacking with 5 base learners + Bayesian hyperparameter tuning	Seasonal datasets (wind & solar)	Outperformed baseline methods with higher R^2 and generalization	Needs frequent retraining; limited flexibility in volatile environments
Khan et al. (2022)	Deep Stacked Ensemble (ANN + LSTM + XGBoost) with SHAP explainability	4 solar energy datasets	R^2 improved by 10–12% over individual models; consistent across weather	Computationally expensive; slower for real-time deployment
Alghamdi et al. (2023)	Stacking ensemble optimized with hybrid GABER algorithm	Renewable energy datasets (wind speed + solar radiation)	High accuracy, stability, and generalization compared to baselines	Algorithmic complexity hinders scalability for real-time
Abdellaif et al. (2022)	Stack-ETR (RF + XGB + AdaBoost → ETR meta-learner)	PV datasets (4 years, 3 PV systems)	RMSE reduced by 40.2%, MAE by 47.2% vs baseline ETR	Requires long-term, high-quality meteorological data
Liao et al. (2024)	SHAP, PFI, PDP, LIME for XAI in wind forecasting	Wind power datasets	Enabled feature importance interpretation; improved trust in models	Interpretation credibility depends on chosen XAI method
Yaseen et al. (2025)	X-Double LSTM + modified SHAP	Wind power time-series (TensorFlow/Keras, Google Colab)	Outperformed LSTM, GRU, Transformer; high accuracy & interpretability	Computationally intensive; limited to specific settings
Kabir et al. (2024)	XAI + symbolic regression for rain effect on HAWTs (BEM theory)	S809 airfoil simulations under rainfall	Predicted aerodynamic changes with R^2 = 0.97; showed 5.7–7%	Simulation-dependent; may not generalize to real turbine

			power loss	operations
Kim & Kim (2023)	ML with enriched meteorological features (turbulence, stability, etc.)	Wind farm power datasets	Improved prediction of 10-min average power	Requires detailed meteorological data; may not be available in all regions

3 Research Methodology:

3.1 Dataset Description

The data provided in this paper is a multi-annual 10minutes interval SCADA (Supervisory Control and Data Acquisition) data measured in a Wind turbine (Vestas V52), which was installed in the Dundalk Institute of Technology, Ireland. It covers a range of 14 years (January 30, 2006 to March 12, 2020) with data being observed at a high frequency of operation permitting an in-depth analysis of the data in a temporal dimension. The wind turbine is a behind-the-meter device in a peri-urban context with hub height of 60 meters as well as a rotor diameter of 52 meters. It is important to note that between October 4, 2018, and July 27, 2019, the turbine was carrying out a gearbox changeout, so there was a lack of positive electricity power production.

Such data consists of 10 minutes readings of different turbine and environmental factors that are crucial in the generation of power. WindSpeed, WindDirAbs, and StdDevWindSpeed use parameters that give the behavior of the wind, whereas Power, MaxPower, MinPower, and AvgRPOW also give the electrical power. Revolution parameters of the mechanical components are monitored through those parameters, such as GenRPM and RotorRPM. EnvirTemp, NacelTemp and various generator and gearbox temperatures such as GearOilTemp, GenTemp, GenBearTemp also monitor environmental and internal conditions. These 19 features will form a strong basis of making predictions.

The data generated by SCADA is continuous and high-resolution and most importantly rich in data making it suitable in machine learning (Ex. time-series forecasting, feature importance). A set of data since 2015 was used to get manageable computational load. As well, the parameters of wind speed are directly measured with a 2 -D ultrasonic Thies Clima anemometer installed in turbine nacelle, which guarantees their high precision. The large extent of the dataset can contribute to predictive modeling as well as employ feasible AI methods like LIME, so they can interpret the behavior of the models on the basis of these physical parameters.

3.2 Data Preprocessing

In data preprocessing of this study, normalization is a very important step to an accurate prediction of wind power using SCADA dataset. The MinMaxScaler of Scikit-learn is used to make sure that any feature contributes equally to the training process: the data is transformed to a predefined range of 0-1. The approach particularly works when dealing with those algorithms, which are sensitive to the scale of the input values, including even gradient-based models and support vector machines. Consequently, due to this transformation the values of power are rescale in a normalized form so as to facilitate easier convergence of models as well as make predictions more stable. Example values of powers after scaling therefore have consecutive values of 0.6422, 0.6376, 0.6354 etc to the 26303rd row. With this transformation and normalization of the data, the study forms a good foundation of consistency of what is being represented as input to the following feature engineering analysis, model training, and perhaps explaining of artificial intelligence through LIME. Such a process will have a direct impact on enhancing model performance and interpretability.

3.3 EDA

An example of a bar chart can be seen in Figure 1 that demonstrates daily average power production during an average month. Each bar denotes average power output on a given day, and the color-coded bars were useful in color contrast as well. It is prudent to mention that typically an average of 204.77 kW was the highest average power generation on Day 14, and a similar average of about 190-200 kW generation was observed on Days 5, 7, 13 and 17. On the other hand, Leap days such as 21 and 22 have lower averages descending towards 160 kW. The visualization will assist in the model training and interpretation of the variability of performance across time as it is relevant to trends and fluctuations in wind power generation occurring daily.

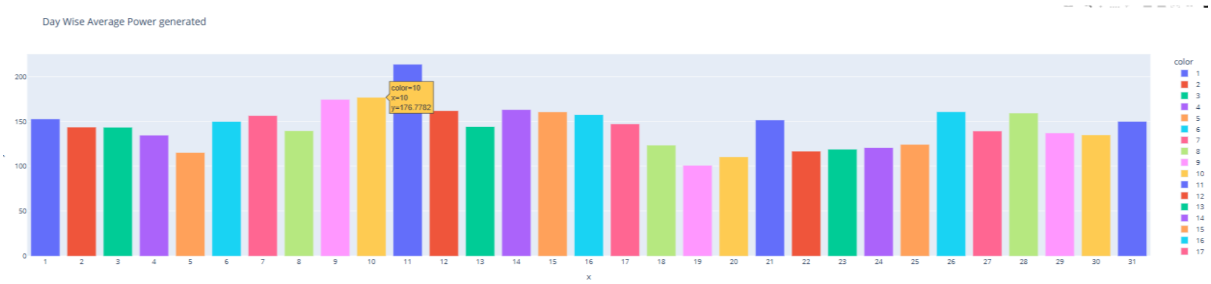


Figure 1: Bar chart

Figure 2 presents a time plot series diagram that displays the fluctuation of the minimum, average, and maximum wind power that was generated between October of 2019 and March of 2020. Blue scale indicates the minimum power, green scale maximum power and red scale shows average power at each timestamp. The maximum values of the power go above 3000 kW around mid-December 2019, and at the beginning of January 2020, and the average in all instances remains between 500 and 1500 kW. There is negative values with sharp dips as outliers or anomny. On the whole, the figure successfully visualizes the dynamics of changes in wind energy production over time, which contributes to the detection of the periods of high variance and guarantees the improved reliability of forecasts.

Figure 2 is a time series plot that shows the fluctuation of the minimum, average, and maximum wind power generated during the period between October of 2019 and March of 2020. The green curve displays the maximum power, red one is average, and blue displays the lower power in all the timestamps. It can be seen that the peaks occur in the middle of December

2019 and in the beginning of the January 2020, with the highest power exceeding 3000 kW, whereas average values remain at the same range of 50031500 kW. There are some sharp downward dips into negative figures that are considered as anomalies or outliers. In general, the figure can be stated as a clear representation of the temporal variations of wind power generation, which will help to predict the high-variance periods as well as promote the superior reliability of forecast.

Figure 2: Time series plot

A bar chart has been provided showing the average yield distribution of wind speed according to month in figure 3. February (Month 2) reports the highest average wind speed (about 7.6 m/s), close behind which are January (Month 1) and March (Month 3) with the wind speeds higher than 7.2 m/s. On the contrary, June (Month 6) and July (Month 7) are the months that show the least average wind speeds of about 6.2 to 6.4 m/s. The rest of the months like October (10), November (11), and December (12) also have moderate wind speed of the level between 6.5 to 6.9 m/s. This monthly pattern gives information about the behavior of seasonal wind, which is also necessary in forecasting power.

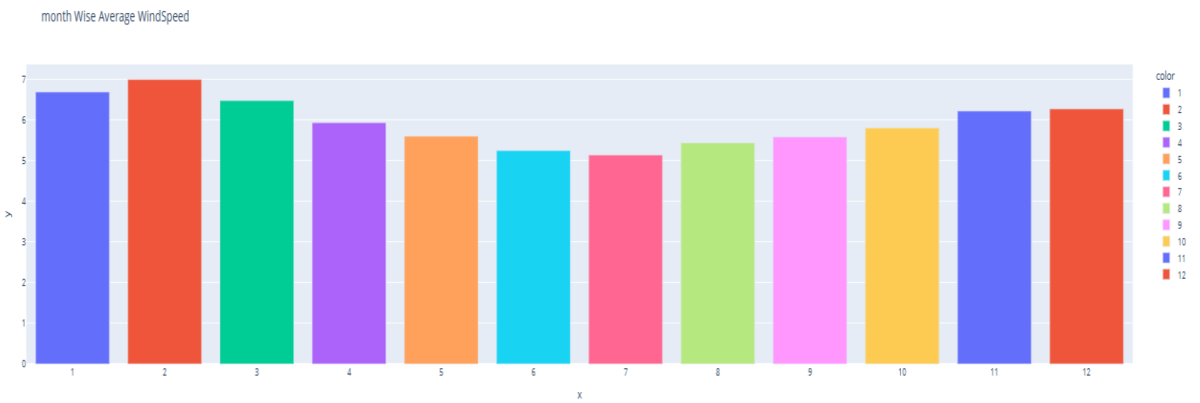


Figure 3: Bar Chart

3.4 Feature Engineering

During the process of Feature Engineering in the given study, one of the most important transformation procedures known as window rolling is utilized to process the time series input data so that the input data is ready to be modeled. In particular, the SCADA dataset is used, and the Power column is processed in rolling window way with a size of 15. This method applies the 15-day running average of the last 15 successive values of power to smooth out the short-term variation and emphasize the longer-term trends or trends on the observed data in time. The outcome is a superior time series that illustrates the wind power generation behavior in a better way.

This rolling average is also called rollingwindowdata, and it helps the model to learn the sequential dependence of one piece of data to the other. It has been observed that the following are the features of the dataset such as Power, MaxPower, and MinPower and the corresponding timestamps are noted as 2019-09-05 to 2020-03-12 containing the approximate 26303 rows of reading. These values are equivalent to the power or kilowatts (kW) within a period of 10 minutes. Using the rolling window, new data point is an average of the last 15 readings, making the model able to generalize more effectively, as well as eliminating sudden spikes or noise. This step of feature engineering becomes central when it comes to getting the data ready in order to build sound forecasting models.

Figure 4 presents line graphs of the differences on the comparison of the actual average power (blue line) and the 15-point window rolling average (red line) of wind turbine output over the period of time. An x axis is presented in the form of timestamps on September 2019- March 2020 and a y axis reflects the values of power in kilowatts (kW). Raw power values are filled with sudden fluctuations both upwards (spikes up to 800 KW) and downwards (negative values sporadically dropping as low as -1000 KW) because of anomalies or other cases of missing data. In comparison, the red line (rolling average) enhances these short-term changes and depicts fundamental trends more visibly. The visual illustrates the ease that the rolling window method will help in the reduction of noise and model preparedness in forecasting.

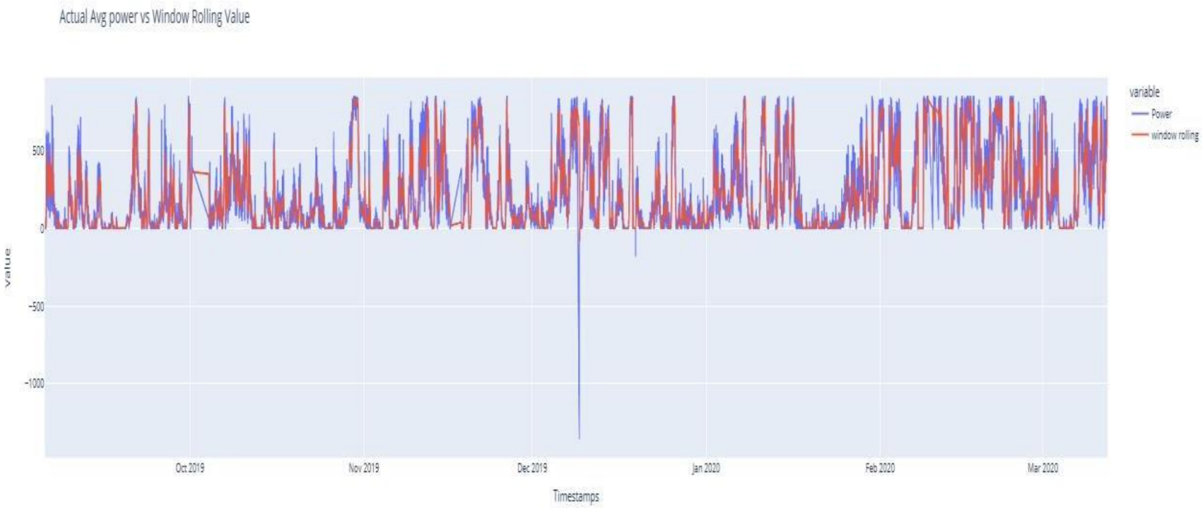


Figure 4: Line Plot

4 Design Specification:

As shown in Figure 5, stacked ensemble learning approach was applied to design the overall system architecture to predict wind power. The process starts with the data source which in this case is the SCADA dataset that was gathered on the Vestas V52 wind turbine. The raw data will be completely preprocessed, i.e. loaded, filtered (by picking the pertinent time periods such as 2015), and normalized (based on the MinMaxScaler). In addition, the exploratory data analysis (EDA) will be applied to know the distribution of data and the correlations between features. In the second step feature engineering is done via rolling and sliding windows to capture time-series behaviour and organize data into forms suitable to supervised learning. Train-test split is used to facilitate the process of performance evaluation. Two machine learning models; SVR, Random Forest, XGBoost, and AdaBoost will be trained to learn different characteristics of the pattern of the data with the processed features being piped into them. A stacked ensemble is made of such base learners: their predictions are obtained and then aggregated by a meta-learner to increase the overall predictive accuracy and robustness. Such modular and layered architecture does not only increase the performance in terms of predictiveness but also help with explainable AI-backed interpretability and establish a scalable basis to execute projects into practice into smart energy systems.

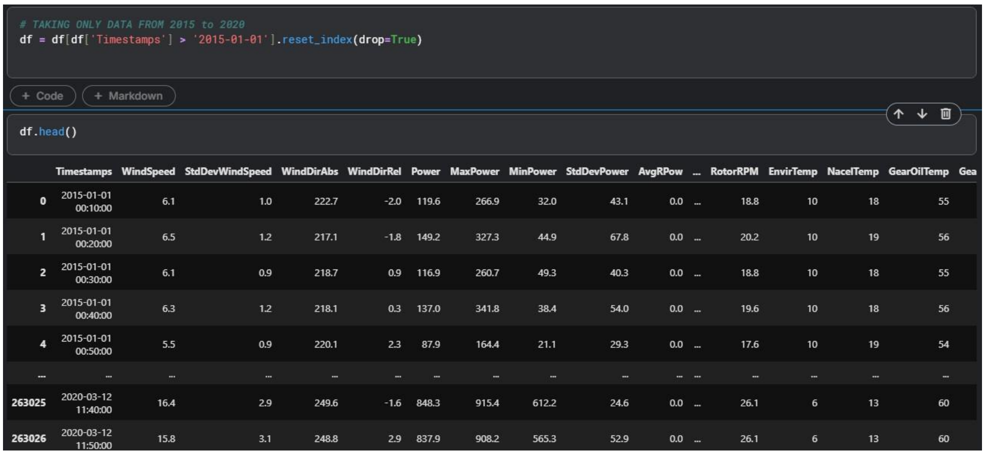
Figure 5: System Architecture Diagram

5 Implementation:

5.1 Data Filtering Based on Resource Constraints

To ensure efficient processing and manage memory constraints, only the records from the year 2015 onwards are selected from the complete SCADA dataset, which spans multiple years.

This filtering significantly reduces the dataset size while retaining meaningful patterns for forecasting. The Timestamp column is used to apply the filter, and the dataset is reset to maintain indexing continuity. This approach enables smoother time-series transformations and model training without overloading the system's computation resources as shown in Figure 6.



```
# TAKING ONLY DATA FROM 2015 to 2020
df = df[df['Timestamps'] > '2015-01-01'].reset_index(drop=True)
```

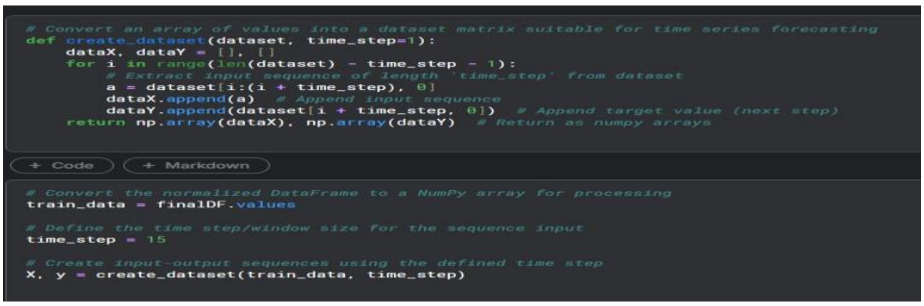
```
df.head()
```

	Timestamps	WindSpeed	StdDevWindSpeed	WindDirAbs	WindDirRel	Power	MaxPower	MinPower	StdDevPower	AvgRPow	...	RotorRPM	EnvirTemp	NacelTemp	GearOilTemp	Gea
0	2015-01-01 00:10:00	6.1	1.0	222.7	-2.0	119.6	266.9	32.0	43.1	0.0	...	18.8	10	18	55	
1	2015-01-01 00:20:00	6.5	1.2	217.1	-1.8	149.2	327.3	44.9	67.8	0.0	...	20.2	10	19	56	
2	2015-01-01 00:30:00	6.1	0.9	218.7	0.9	116.9	260.7	49.3	40.3	0.0	...	18.8	10	18	55	
3	2015-01-01 00:40:00	6.3	1.2	218.1	0.3	137.0	341.8	38.4	54.0	0.0	...	19.6	10	18	56	
4	2015-01-01 00:50:00	5.5	0.9	220.1	2.3	87.9	164.4	21.1	29.3	0.0	...	17.6	10	19	54	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
263025	2020-03-12 11:40:00	16.4	2.9	249.6	-1.6	848.3	915.4	612.2	24.6	0.0	...	26.1	6	13	60	
263026	2020-03-12 11:50:00	15.8	3.1	248.8	2.9	837.9	908.2	565.3	52.9	0.0	...	26.1	6	13	60	

Figure 6: Data Filtering Output (Table View)

5.2 Time Series Input Preparation using Sliding Window

In the implementation of this time series prediction framework, the goal is to filter all the other features so that only the Power column of the SCADA dataset is implemented. This is to reduce the problem to univariate forecasting task as indicated in Figure 7. The data is processed to be ready to feed to machine learning models by employing the sliding window process. The window size is set to 15, which implies that the model will rely on the previous 15 time steps to find out the value of the upcoming level of power out. The sequence data would here be converted into supervised learning that has distinct input-output pairs. The implementation of code specifies a bespoke function create_dataset() which reads over the normalized power data and generates training samples. The samples are made up of successions of 15 successive values of power used as training, with the 16th value of power used as the predictive target. It returns two NumPy arrays X and y which represent model inputs and model outputs, respectively. It is a process by which the ML models can learn the temporal dependencies showing in the historic data efficiently. A combination of fixed-length time step allows the patterns of wind turbine generating capacity to be captured to the extent that the model can be applied to predict future values successfully based on historical patterns.



```
# Convert an array of values into a dataset matrix suitable for time series forecasting
def create_dataset(dataset, time_step=1):
    dataX, dataY = [], []
    for i in range(len(dataset) - time_step - 1):
        # Extract input sequence of length 'time_step' from dataset
        a = dataset[i:(i + time_step), 0]
        dataX.append(a)  # Append input sequence
        dataY.append(dataset[i + time_step, 0])  # Append target value (next step)
    return np.array(dataX), np.array(dataY)  # Return as numpy arrays
```

```
# Convert the normalized DataFrame to a NumPy array for processing
train_data = finalDF.values

# Define the time step/window size for the sequence input
time_step = 15

# Create input-output sequences using the defined time step
X, y = create_dataset(train_data, time_step)
```

Figure 7: Sliding Window Time Series Sequence Generation

5.3 Model Explainability Using LIME

In case of the stacked ensemble model it was necessary to interpret their predictions, and LIME (Local Interpretable Model-agnostic Explanations) was applied. LIME is used to find the contribution of each of the input features (here the values of lags measured in time series) to individual prediction. The figure below indicates that characteristics such as lag 14, lag 12 and lag 5 were effective in a positive way in creating positive result in the prediction output as illustrated in Figure 8. As is reflected in the figure, all other characteristics exerted a less positive or none positive influence in the prediction output. The visual representation emphasizes openness in the decision-making of the model, which is an essential point in the energy systems. These explanations allow not only the process of making forecasts to be accurate but also interpretable.

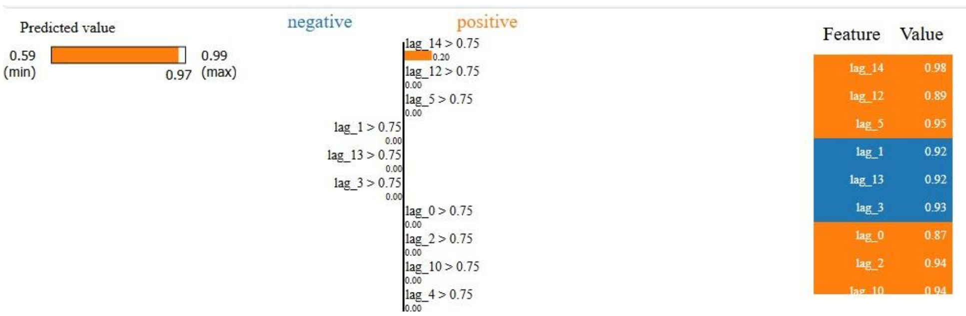


Figure 8: LIME Explanation for Single Prediction Instance

The figure 9 shows the forecasting outcome of the successive 15 time intervals in the stacked ensemble model. The blue line demonstrates the previous 15 actual power generation data used in SCADA data measured on March 12, 2020, whereas red line shows the projected 15 time points. The forecast has a comparatively persistent direction, which consists of the downward fluctuation after noon. The graph confirms the consistency of the model in predicting the future power generation with a reasonable level of accuracy, even on short- term aberrations such as reduction in power generation with time 11:00AM. Such a time series forecasting helps make proactive energy planning and scheduling.

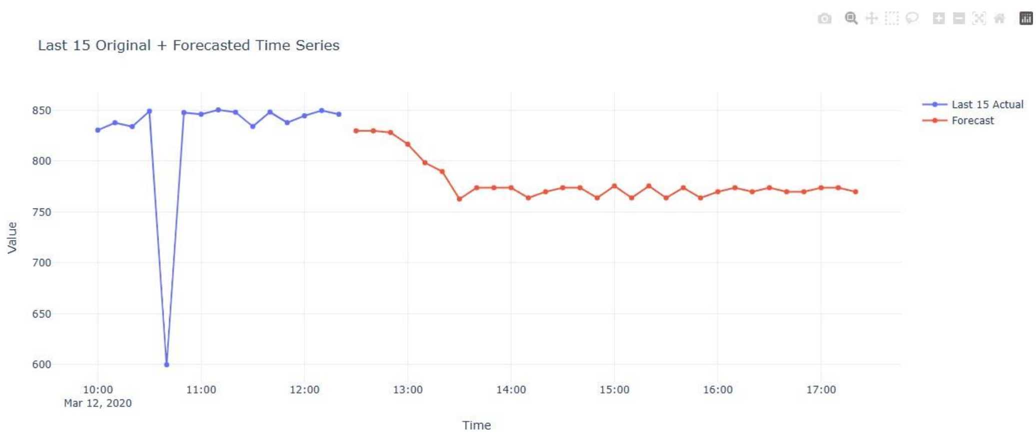


Figure 9: Forecasting Future Power Generation (Line Chart)

6 Evaluation:

6.1 Case Study 1: SVM

The graph of Figure 10 shows the real and estimated powers of wind in the Support Vector Regressor (SVR) model established between February 1 and March 4, 2020. The blue line denotes the SCADA-measured values, and the red line gives the values of the predictions by the SVR model. The nonlinear variation of power is well registered with the SVR, whereby the maxima (~900 kW) and the minimum (~100 kW) are traced closely. Other deviations are witnessed when there are sudden falls of signals or inconsistencies in the sensors. On the whole, the prediction trend is quite in agreement with the real-time variations, which proves the SVR use in describing complicated trends in wind turbine power generation.

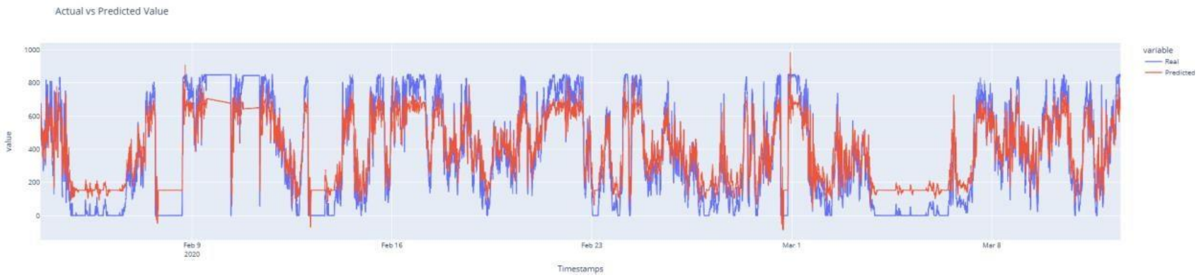


Figure 10: Actual vs Predicted Value

6.2 Case Study 2: Adaboost

Figure 11 shows the comparison of actually observed wind power and the wind power predicted via the AdaBoost Regressor during February 1- March 4, 2020. The blue line is the real SCADA power output and the red line is what AdaBoost model actually predicted. The model is equally accurate in describing the overall power trend in the range of 0 to 950 kW wherein it is well represented particularly during stable operational periods. Although there are small misalignments in sudden (dips) declines (e.g. February 10 -12), AdaBoost boosting process aids to fine-tune low-powered learners to increase the accuracy in the erratic power cycles and eruptive series.

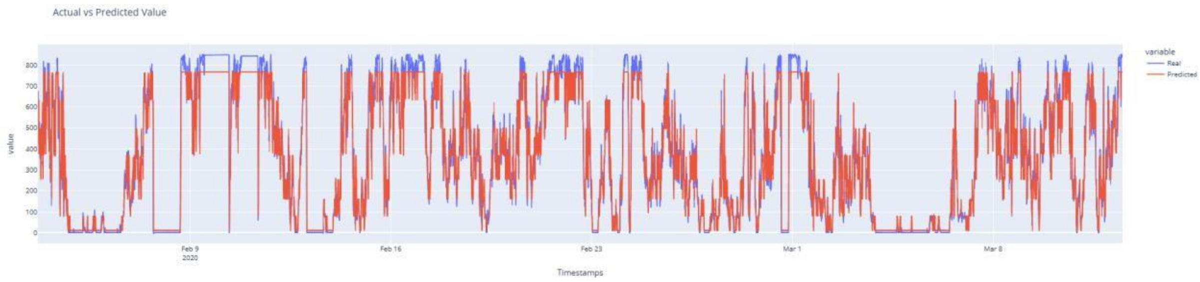


Figure 11: Actual vs Predicted Value

6.3 Case Study 3: Random Forest

Figure 12 shows the actual and predicted (mean) wind turbine power output over the period February 1 to March 4, 2020 by using the Random Forest Regressor. The blue line refers to the real values recorded being SCADA, and the red line refers to the estimated values provided by the Random Forest model. Pronounced maximum values of the model up to ~950 kW allow the model to reflect the real output in both constant and variable periods. The high correlation of high-density areas and low decay in transition indicates that Random Forest has the advantage of being able to capture non-linear relationships between the variables and allow them to be represented through several decision trees, which allows highly reliable and faithful forecasting by this method.

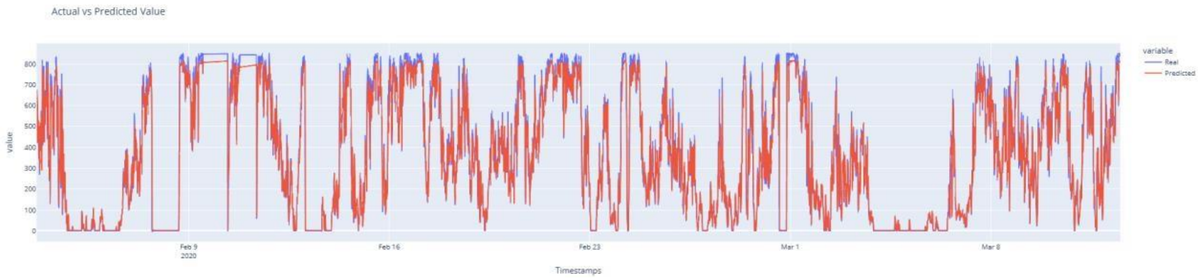


Figure 12: Actual vs Predicted Value

6.4 Case Study 4: XGBoost

Figure 13- The real and estimated wind power output with the XGBoost Regressor across the dates of 1-2-20 to the 4 of March 2020. The red curve is the XGBoost and blue curve is the SCADA-measured values of power. The model convincingly follows the variations in power ranging between 100 kW and 950 kW during fast variation as well as high-frequency cycles. The gradient boosting model of XGBoost allows it to deal with nonlinear relationships and noise and that can be seen in the fact that XGBoost is very close to the true values at the majority of intervals. This confirms its appropriateness in strong energy forecasting work, high-precision.

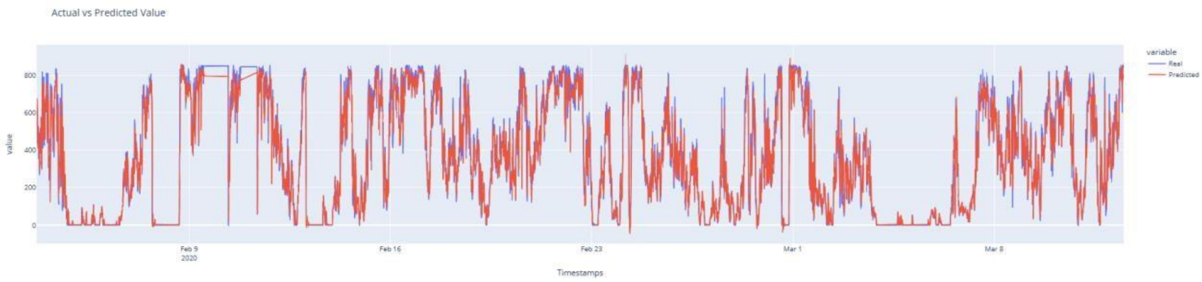


Figure 13: Actual vs Predicted Value

6.5 Case Study 5: Stacking

The comparison between the actual and predicted wind power output with the help of the proposed Stacking Ensemble Model shows the results of the periods over February 10 to March 4, 2020 (Figure 14). The blue line shows the real measured values that are recorded by

SCADA and the red line shows the actual values predicted by the ensemble model. The ensemble strategy, whereby the results of SVR, Random Forest, and XGBoost, AdaBoost are combined and used as a forecast, demonstrates superior levels of prediction with an outstanding match between 100 kW and 950 kW. It reduces prediction errors in spite of noisy transitions, which is superior to individual models in consistent and generalizable prediction. This confirms the effectiveness of the ensemble when it comes to predictive power of robust winds.

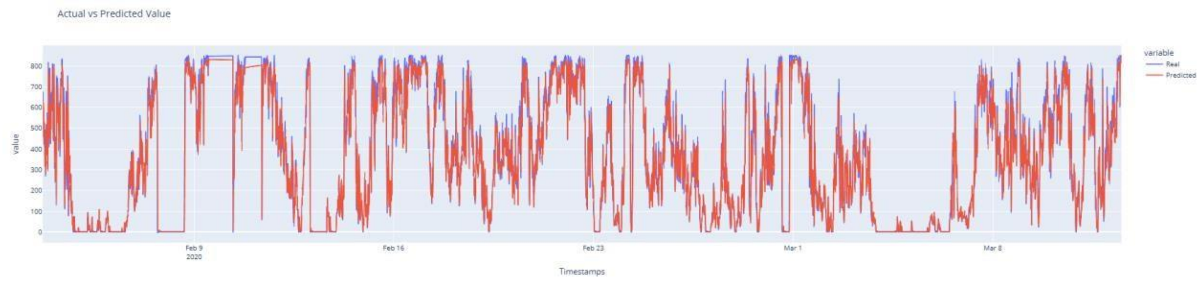


Figure 14: Actual vs Predicted Value

Table 1 constitutes a detailed comparison of five machine learning models of prediction focus on wind power through the use of SCADA data, measured in Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R 2 Score. The Random Forest Regressor had the best performance and it performed better, obtaining an R 2 of 0.9251 with minimum Root Mean Square Error of 0.0358, which effectively fits the non-linear trends in power generation. The suggested Stacking Ensemble Model, however, that incorporates predictions of SVR, Random Forest, AdaBoost, as well as XGBoost, possessed an even higher overall reliability with the R 2 equal to 0.9231. In order to confirm whether or not it generated generalizable performance, we used cross-validation or 5-fold cross-validation on the stacking ensemble. The cross-validated results also enhanced the accuracy whereby it had an average MSE of zero point 0009, 0.0293 RMSE and 0.9292 R 2 and therefore it was the most accurate model in the study. These findings underscore the unprecedentedness of combining ensemble learning and careful validation giving rise to both robustness and predictive accuracy. The stacking method becomes an efficient solution to the individual bias of the models and it is a scalable and reliable while one time forecasting wind power in real-world scenarios.

Table 2. Performance Comparison of ML Models for Wind Power Forecasting

Model	(MSE)	(RMSE)	R^2 Score
Support Vector Regressor (SVR)	0.00318	0.0564	0.8133
AdaBoost Regressor	0.00145	0.0381	0.9148
Random Forest Regressor	0.00128	0.0358	0.9251
XGBoost Regressor	0.00135	0.0368	0.9208
Stacking Ensemble Model	0.00131	0.0362	0.9231
Stacking (Cross-Validated)	**0.0009**	**0.0293**	**0.9292**

6.6 Comparative Evaluation with Related Study

To further validate the applicability of stacked ensemble models in renewable energy forecasting, we compared our proposed stacking model with the Stack-ETR framework developed by Abdellaif et al. (2022). Although their study focused on photovoltaic (PV) power output, the methodological similarities allow for a meaningful comparison as shown in Table 3.

Table No: 3

Study	Model	Task	R^2	RMSE	MAE
My proposed work	Stacking (SVR + RF + XGB + AdaBoost, 5-fold CV)	Wind power forecasting using SCADA	**0.9292**	**0.0293**	–
Abdellaif et al. (2022)	Stack-ETR (RF + XGB + AdaBoost → ETR)	PV power forecasting (1-day-ahead)	–	Significant reduction: RMSE improved by 40.2% compared to standalone ETR	MAE improved by 47.2% compared to standalone ETR

Both studies highlight the performance of stacking in improving predictive performance over standalone learners. While Abdellaif et al. (2022) showed significant error reduction in PV forecasting, our model achieved higher R^2 (0.9292) and lower RMSE (0.0293) in wind power forecasting with SCADA data. Additionally, our integration of LIME-based interpretability provides transparency, making the model suitable for real-time smart grid applications.

7 Conclusion and Future Work:

This study was able to demonstrate a successful use of machine learning models which is a stacked ensemble model to predict the power of wind turbines using SCADA data. When comparing the performance of the four models evaluated (SVR, Random Forest, AdaBoost, and XGBoost), the ensemble model demonstrated the best quality, having taken the advantage of the capabilities of separate learners. The accuracy of the stacking model proposed and tested with 5-fold cross-validation was amazing based on a score achieved of 0.9292 showing the R2 scores. Besides, the explainable AI methods, in particular, LIME, implemented, made model predictions transparent, discovering the important features of SCADA that affect power outputs. These two features of predictive accuracy and explainability promote the feasible usage in real-time energy systems. Nevertheless, there are still a few issues, such as the noise sometimes may appear in signals, and it cannot be computed within big data size. In future work, we would look into incorporating deep learning models including LSTM and Transformer-based architectures to reflect longer dependencies and the dynamics of time in a better manner. To increase reliability of forecasts, inclusion of real-time weather inputs and external grid inputs can also be included. In addition, the scalability on smart grid systems and to be able to operate on the cloud platforms with autoscaling capability can be enforced by deployment of the framework on the cloud platforms. On the whole, the upcoming study adds a well-performing, explainable, and scalable solution to the problem of renewable energy forecasting as it involves both technical and operational aspects.

References:

[1] Ikesaiberi, A., Harrou, F. and Sun, Y., 2022. Efficient wind power prediction using machine learning methods: A comparative study. *Energies*, *15*(7), p.2327.

[2] Tarek, Z., Shams, M.Y., Elshewey, A.M., El-kenawy, E.S.M., Ibrahim, A., Abdelhamid, A.A. and El-dosuky, M.A., 2023. Wind Power Prediction Based on Machine Learning and Deep Learning Models. *Computers, Materials & Continua*, *75*(1).

[3] Karaman, Ö.A., 2023. Prediction of wind power with machine learning models. *Applied Sciences*, *13*(20), p.11455.

[4] He, R., Yang, H., Sun, S., Lu, L., Sun, H. and Gao, X., 2022. A machine learning-based fatigue loads and power prediction method for wind turbines under yaw control. *Applied Energy*, *326*, p.120013.

[5] Dhiman, H.S . and Deb, D., 2020. A review of wind speed and wind power forecasting techniques. *arXiv preprint arXiv:2009.02279*.

[6] Hanifi, S., Liu, X., Lin, Z. and Lotfian, S., 2020. A critical review of wind power forecasting methods—past, present and future. *Energies*, *13*(15), p.3764.

[7] Huang, H., Zhu, Q., Zhu, X. and Zhang, J., 2023. An adaptive, data-driven stacking ensemble learning framework for the short-term forecasting of renewable energy generation. *Energies*, *16*(4), p.1963.

[8] Khan, W., Walker, S. and Zeiler, W., 2022. Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. *Energy*, *240*, p.122812.

[9] Alghamdi, A.A., Ibrahim, A., El-Kenawy, E.S.M. and Abdelhamid, A.A., 2023. Renewable energy forecasting based on stacking ensemble model and Al-Biruni earth radius optimization algorithm. *Energies*, *16*(3), p.1370.

[10] Abdellatif, A., Mubarak, H., Ahmad, S., Ahmed, T., Shafiullah, G.M., Hammoudeh, A., Abdellatef, H., Rahman, M.M. and Gheni, H.M., 2022. Forecasting photovoltaic power generation with a stacking ensemble model. *Sustainability*, *14*(17), p.11083.

[11] Kamani, D. and Ardehali, M.M., 2023. Long-term forecast of electrical energy consumption with considerations for solar and wind energy sources. *Energy*, *268*, p.126617.

[12] Czapaj, R., Kamiński, J. and Sołtysik, M., 2022. A review of auto-regressive methods applications to short-term demand forecasting in power systems. *Energies*, *15*(18), p.6729.

[13] Udo, W.S., Kwakye, J.M., Ekechukwu, D.E. and Ogundipe, O.B., 2024. Optimizing wind energy systems using machine learning for predictive maintenance and efficiency enhancement. *Journal of Renewable Energy Technology*, *28*(3), pp.312-330.

[14] Chen, S., Zhang, C. and Lu, X., 2023. Energy conversion from Fossil fuel to renewable energy. In *Handbook of Air Quality and Climate Change* (pp. 1-44). Singapore: Springer Nature Singapore.

[15] Munoz, M.A., Morales, J.M. and Pineda, S., 2020. Feature-driven improvement of renewable energy forecasting and trading. *IEEE Transactions on Power Systems*, *35*(5), pp.3753-3763.

[16] Sarker, I.H., 2021. Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), p.160.

[17] Liao, W., Fang, J., Ye, L., Bak-Jensen, B., Yang, Z. and Porte-Agel, F., 2024. Can we trust explainable artificial intelligence in wind power forecasting?. *Applied Energy*, *376*, p.124273.

[18] Yassen, M.A., El-Kenawy, E.S.M., Abdel-Fattah, M.G., Ismail, I. and Mostafa, H.E.D.S., 2025. Explainable artificial intelligence for wind power forecasting model based on long short-term memory. *Neural Computing and Applications*, pp.1-23.

[19] Syed Ahmed Kabir, I.F., Gajendran, M.K., Taslim, P.M.P., Boopathy, S.R., Ng, E.Y.K. and Mehdizadeh, A., 2024. An XAI framework for predicting wind turbine power under rainy conditions developed using CFD simulations. *Atmosphere*, *15*(8), p.929.

[20] Kim, D.Y. and Kim, B.S., 2023. Contribution of meteorological factors based on explainable artificial intelligence in predicting wind farm power production using machine learning algorithms. *Journal of Renewable and Sustainable Energy*, *15*(1).