ANALYSIS OF FACTORS AFFECTING WIND TURBINE ENERGY OUTPUT USING MACHINE LEARNING

By

GHULAM MURTAZA

NATIONAL UNIVERSITY OF MODERN LANGUAGES, ISLAMABAD

January 2024

ANALYSIS OF FACTORS AFFECTING WIND TURBINE ENERGY OUTPUT USING MACHINE LEARNING

By

GHULAM MURTAZA

B.Sc. Electronic Engineering, The Islamia University of Bahawalpur, 2015

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

Electrical Engineering

TO

FACULTY OF ENGINEERING AND COMPUTING

NATIONAL UNIVERSITY OF MODERN LANGUAGES ISLAMABAD

THESIS AND DEFENCE APPROVAL FORM

The undersigned certify that they have read the following thesis, examined the defense, are satisfied with overall exam performance, and recommend the thesis to the Faculty of Engineering and Computer Science for acceptance.

Thesis Title: Analysis of Factors Affecting Wind Turbine Energy Output Using Machine Learning

Submitted By**:** Ghulam Murtaza Registration #: MS/EE/004

Master of Science in Electrical Engineering

Electrical Engineering
Discipline

Dr. Sajid Saleem _____
Research Supervisor Signature of Research Supervisor

Dr. Noman Malik _____
Dean (FEC) Signature of Dean (FEC)

Brig. Shahzad Munir _____
Director General Signature of Director General

January 30th , 2024
Date

AUTHOR'S DECLARATION

I **Ghulam Murtaza**

Son of **Ghulam Hussain**

Registration # **004/MS/EE/F20**

Discipline **Electrical Engineering**

Candidate of **Master of Science in Electrical Engineering (MSEE)** at the National University of Modern Languages do hereby declare that the thesis **Analysis of Factors Affecting Wind Turbine Energy Output Using Machine Learning** submitted by me in partial fulfillment of MSEE degree, is my original work, and has not been submitted or published earlier. I also solemnly declare that it shall not, in the future, be submitted by me for obtaining any other degree from this or any other university or institution. I also understand that if evidence of plagiarism is found in my thesis/dissertation at any stage, even after the award of a degree, the work may be canceled, and the degree revoked.

_____

Signature of Candidate

Ghulam Murtaza

Name of Candidate

30 JANUARY 2024

Date

ABSTRACT

Analysis of factors affecting wind turbine energy output using machine learning

Electrical energy generated by wind turbines is stochastic in nature due to its dependency on various factors. Such randomness raises barriers in adjusting the energy stocks of the power systems according to need. Multiple approaches have been proposed to predict the energy output of wind turbines and to meet the corresponding energy demands. This thesis investigates variables (also called factors or features) that affect the wind turbine's output. The energy obtained from turbines varies as it relies on factors. Some of the important factors or features are turbine blade area, wind speed, temperature, air density, humidity, tower height, the angular position of the blades, pressure, etc. All such features are required to be investigated, analyzed, and evaluated with the help of state-of-the-art Machine Learning (ML) models to identify their importance and significance in predicting the wind turbine output. ML techniques used are Categorical Boosting (CatBoost), Extreme Gradient Boosting (XGBoost), Decision Tree (DT), Random Forest (RF), Gradient Boosting Regression (GBR), Light Gradient Boosting Model (LGBM), Extra Tree and Adaptive Boosting (AdaBoost). Evaluation of the ML methods and analysis of the factors are carried out on three different latest and publically available datasets. The experimental results show that CatBoost compared to all other methods demonstrates the lowest Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) in predicting the energy output. The wind speed is identified as the most significant factor in predicting the energy output by all of the ML methods. Additionally, a new method is proposed which ensembles CatBoost, RF and AdaBoost methods. The proposed method is named as X-CRA, in which the prediction s of CatBoost, RF and AdaBoost are fed into XGBoost through a stacking approach and final energy output is obtained. The experimental results show that X-CRA outperforms CatBoost and all other ML methods in predicting the wind energy output.

TABLE OF CONTENT

List of Tables

TABLE NO. **TITLE** **PAGE**

List of Figures

FIGURE NO. **TITLE** **PAGE**

LIST OF ABBREVIATIONS

ANN	Artificial Neural Networks
DT	Decision Tree
GBM	Gradient Boosting Machine
GBR	Gradient Boosting Regression
KNN	K-Nearest Neighbors
LGBM	Light Gradient Boosting Machine
LS-SVM	Least-Squares Support Vector Machine
LSTM	Long-Short Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Square Error
RF	Random Forest
RMSE	Root Mean Square Error
SCADA	Supervisory Control And Data Acquisition
SVM	Support Vector Machines
WP	Wind Power
XGBoost	Extreme Gradient Boosting

ACKNOWLEDGEMENT

First and foremost, I am deeply grateful to Allah Almighty for bestowing His blessings upon me. I would like to extend my heartfelt appreciation to my parents for their unwavering support throughout my research journey. Their love, encouragement, and unwavering belief in my abilities have been instrumental in my accomplishments.

I am also indebted to my research supervisor, Dr. Sajid Saleem, for his persistent dedication and invaluable guidance in completing this research. His unwavering enthusiasm and expertise in the field of research have continuously inspired me. I am truly grateful for his invaluable advice, comprehensive support, and unwavering involvement in every aspect of my research. His exceptional guidance has played a crucial role in shaping the outcome of this thesis.

Once again, I extend my heartfelt gratitude to Allah Almighty, my parents, my supervisor and my friend Hamza Abbasi for their invaluable contributions to my academic journey. Their guidance, support, and friendship have been invaluable, and I am truly grateful for their presence in my life.

DEDICATION

Thesis is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

CHAPTER 1

INTRODUCTION

As global energy requirements continue to increase, researchers are increasingly focusing on developing innovative and effective methods for energy generation [1]. Energy has become a basic component nowadays and efforts are being made to find efficient and renewable energy sources. Energy obtained from sources are required to be environment-friendly and have potential for maximum energy output at a particular time and environment. Energy sources which are environmentally friendly and low generation cost are investigated and introduced at priority globally and seen as a replacement for carbon-based energy [2].

1.1 Wind Energy

Amongst renewable energy sources, the wind turbines have emerged as a significant and eco-friendly means of electricity generation [3], [4]. Electricity generated by wind turbines is clean and environmentally friendly and can be generated at a low cost. Windmills became widespread in Europe in the 12th century and were used for a variety of purposes, including grinding grain, pumping water, and sawing wood [5]. First wind turbine for electricity generation was built in 1887. Turbine was used to power the lighting in a cottage in Scotland [6]. In 1891, Poul la Cour built a wind turbine that was used to generate electricity for a blacksmith's shop [7].

In the early 20th century, wind turbines were used extensively in rural areas of the United States to generate electricity. These turbines were small and used to power individual homes or farms. In response to the energy crises of the 1970s, large-scale wind turbine development was initiated [7]. These advanced turbines revolutionized electricity generation offers greater efficiency and the ability to generate power on a larger scale compared to earlier sources.

Presently, the growth of wind energy is outpacing many other renewable energy sources. In 2020, according to the International Energy Agency, wind energy will account for 7.1% of

the world's electricity generation, up from 2.6% in 2010 [7]. Countries that produce the most wind energy are China, the United States, Germany, India, and Spain.

Prospects of wind energy are bright, given its rapid expansion as a renewable energy source. Several noteworthy trends and projections have been observed [8]. Wind energy capacity is anticipated to grow at an annual rate of 20% until 2025. By 2050, it is projected to contribute approximately one-third of the global electricity supply. Wind energy is becoming more cost-effective and competitive with fossil fuels due to advances in wind turbine technology, such as larger and more efficient turbines. Expansion of offshore wind farms has unleashed the wind energy potential in regions with limited land resources.

Over the past decade, the price of wind energy has consistently decreased, making it more affordable for consumers and businesses. Trend is anticipated to continue as technology advances and economies of scale are realized. Globally, governments are instituting policies to promote renewable energy, such as wind power. This includes tax credits, subsidies, and mandates requiring utilities to obtain a certain portion of their energy from renewable sources. Wind energy produces zero emissions and does not contribute to climate change, making it an attractive option with respect to environment. As more people become aware of the benefits of clean energy, demand for wind energy is anticipated to increase steadily in the future.

Wind energy's future is promising, as it is a growing source of clean, renewable energy that is becoming more cost-effective and competitive with fossil fuels. Amount of energy a wind turbine can produce depends on a number of factors, including the turbine's size, the wind's speed and turbine's motor torque etc.Wind turbines have the potential to play a significant role in Pakistan's energy balance, particularly in regions characterized by high wind speeds, such as the coastal regions of Sindh and Balochistan provinces, where the demand for wind power is particularly high. According to Pakistan's Alternative Energy Development Board, Pakistan Potential of wind power is over 50,000 MW, which is many times greater than its current installed capacity of approximately 2,000 MW [9].

In Pakistan, various policies have been placed and initiatives have been taken to actively encourage the advancement of wind power within the country, including the Alternative and Renewable Energy Policy 2019. The policy aims to increase the share of green energy in the

nation's electricity mix to 30 percent by 2030. Wind energy has the potential to provide Pakistan with a significant source of renewable energy and reduce its reliance on imported fossil fuels. However, there are still a number of obstacles to overcome, including the high up-front costs of wind power projects and the requirement for a more robust infrastructure to support the installation of wind energy into the national grid [1].

Offshore wind turbines and aerial wind turbines are the most popular types of energy harvesting methods. Their energy output depends on where they are installed, climate, the wind speed, humidity, temperature and altitude. Collectively, these parameters determine the energy output of a wind turbine at a specific location. Numerous sites suitable for the installation of wind turbines have been identified around the world, enabling the harvesting of wind energy to accomplish energy needs [10].

1.2 Prediction of Wind Energy

Generation of energy by wind turbines is stochastic by nature, dependent on numerous factors and variables [11]. These factors are natural and also change over time. Amount of these factors determines how much energy can be produced. These factors could not be artificially and humanly controlled nor could be determined in advance. Such randomness raises barriers to adjusting the energy stocks in the power systems according to need and creates problems that become more difficult to solve [12]. Multiple methods have been adopted to find out the wind turbine's energy generation and address the corresponding power demand. Machine Learning (ML) methods for inferring the wind turbine's output are emerging and attracting the researchers' attention [13].

As ML methods efficiently handle the variability of the factors or features, it can accurately predict the amount of energy produced by wind turbines [14]. In this way, the energy output from the wind turbines can be related to the factors influencing it and can be predicted with ML methods. Turbine blade area, wind speed, temperature, air density, humidity, tower height, angular position of blades, etc. are some of the factors that influence the output of wind turbines [15]. These factors are natural and cannot be changed but based on their quantity the energy production can be estimated. All such factors or features are required to be investigated,

analyzed, and evaluated with the help of the modern ML models to identify their significance in predicting the wind turbine output [15].

All the evaluation and analysis of the features are required to be done on different latest datasets, especially on publically available datasets. Such datasets allow the examination of factors of different regions around the globe [16]. predicting of energy is a regression problem where the relationship among the independent variables or factors is established to infer the output dependent variable i.e. electrical energy [17]. This allows the identification and analysis of the variables (factors) that are important in predicting the output energy accurately[18].

1.3 Problem Statement

Electrical energy generated by wind turbines is stochastic. Energy prediction based on factors that affect the generated energy through ML is an active research area. Various methods have been developed to predict the electrical energy produced by wind turbines [19]. With the advent of new technology and sensors new factors are being measured [20], which need analysis with state of the art ML techniques. Also, there is a need to analyze all such factors on the latest publically available datasets so that the energy of wind turbines can be predicted accurately.

1.4 Research Questions

The research question are as follows.

RQ1: Which ML models are best for the prediction of wind energy output?

RQ2: What are the factors that affect the wind energy output and their analysis using ML?

1.5 Objectives

Main objectives of the research are as follows.

- To identify ML models for efficient wind energy forecasting
- To identify the factors that affect the wind energy output and their analysis using ML.

1.6 Scope of Study

Research aims to investigate and analyze the factors that influence the output of wind turbines, while also developing predicting models by employing machine learning techniques. This thesis focuses on factors such as wind speed, temperature, air density, wind direction, and turbine specifications, as well as their impact on wind turbine energy production. The collected data associated with these factors undergoes processing through ML based prediction and analysis. ML methods used are regression models due to the continuous nature of the wind turbine output variable. Research includes performance evaluation of the developed models using performance metrics and a comparative analysis using mean squared error, root mean squared error, MAE, and correlation coefficients.

1.7 Contribution and Significance

Study makes significant contributions to the fields of renewable energy and wind power generation which are as follows.

1.7.1 Understanding of Wind Turbine Output

The thesis contributes to understanding how variables such as wind speed, temperature, air density, wind direction, and turbine characteristics affect the performance of wind turbines by conducting a comprehensive analysis of these factors. Information can help to optimize wind farm design, operational strategies, and maintenance planning.

1.7.2 Enhanced prediction Accuracy

Development of predicting models using machine learning techniques enables more accurate prediction s of wind turbine output. By incorporating multiple factors and utilizing advanced algorithms, the models can capture complex relationships and patterns that traditional statistical methods may overlook. Improved prediction accuracy facilitates better energy production planning, grid integration, and resource allocation in wind power systems. Thesis paves the way for future research and advancement in wind turbine output analysis and forecasting.

The potential of machine learning techniques in enhancing prediction accuracy while also identifying areas that require further investigation and refinement is carried out. It supports ongoing research efforts aimed at refining models, incorporating additional variables, and developing advanced algorithms for better prediction s and optimization of wind energy production. Analysis of features and prediction wind turbine output is a significant topic with far-reaching implications for renewable energy integration, wind farm operation, and the transition towards sustainable energy systems.

1.8 Organization of Thesis

Organization of the thesis is as follows Chapter 1 presents introduction, background, research questions and contributions of the thesis. Chapter 2 presents an overview of relevant literature on wind turbine output analysis, forecasting, and machine learning techniques.

Identification of factors influencing the wind turbine performance and output is presented. Discussion of existing methodologies and models used in wind turbine output analysis and prediction is carried out. Chapter 3 presents Methodology which includes description of the datasets, preprocessing steps, ML methods, proposed method and details of performance evaluation metrics for comparative analysis. Chapter 4 presents experimental results with discussion. Finally, Chapter 5 concludes the thesis with future directions.

CHAPTER 2

LITERATURE REVIEW

Chapter presents a review of relevant literature on wind turbine output analysis, forecasting, and machine learning techniques. Including the identification of factors influencing wind turbine performance and output. Discussion of existing methodologies and models used in wind turbine output analysis and forecasting.

2.1 Overview of Machine Learning Methods

K-Nearest Neighbors (KNN) is used for prediction of wind energy and it is compared with other models such as Random Forest (RF), DT , Extra Tree and Gradient Boosting Machine (GBM) [21]. Deployment of these models seeks to improve the short-term wind energy prediction for a Turkish farm in particular. Variables include speed, direction Temperature are analyzed, and possible results of the prediction are determined.

In [22], use Logistic regression for wind energy forecasting. Regression model predicts energy generation from wind-turbine. Adrian Stetco at el. use machine-learning models for prediction [23]. They employ Naïve Bayes model on SCADA (Supervisory Control And Data Acquisition) data [24]. Two datasets and four appropriate reference models are constructed for comparison tests to validate the proposed model's prediction performance. Suggested model's superiority is proved by experimental findings, which indicate that the proposed model can provide adequate wind power intervals with high confidence and quality.

Mohammed Gendeel et al investigate wind farm data to predict the output energy with precision [25]. Prediction is made using the Least Squares Support Vector Machine (LS-SVM) model. Experimental results show the higher accuracy of the LS-SVM model, indicating its precision in predicting the output energy generated. Prediction of wind energy in the region of Kolkata is based on meteorological data. Several models are employed, and experimental

results indicate that the RF and Decision Tree methodologies exhibit outstanding prediction accuracy, as exemplified by their low MAE [26].

In [27], RF, GBR, and XGB models are compared. It is observed that these models outperform the SVM model, further affirming their superior predictive capabilities. Furthermore, for carrying wind power forecasts, For training patterns, the closest set of patterns calculating the Distance is chosen, and a block-wise training and prediction method is applied [28]. Findings indicate that the persistence technique provides a significant improvement [29]. Compare multiple wind energy prediction algorithms and show that the RF with and without random input produce equivalent prediction s to SVM [30].

Rober Mamani et al. examines the applicability of MERRA-2 satellite datasets and Weather Research and Prediction (WRF) simulations for wind energy assessment in a variety of regions [31]. Annual wind averages and features are discovered through the examination of 41 years' worth of hourly wind velocity information gathered by MERRA-2. Additionally, wind shear and fluxes across Bolivia are examined using WRF models for select months. Key findings reveal variances in the wind speed index from 0.90 to 1.09 across different regions, and they also specific times when wind speeds are at their highest, such as May to October in regions. However, the paper notes differences between MERRA-2 data and WRF prediction s and explains them as being caused by site topography.

Irregular and unpredictable nature of Wind Power (WP) hinders the successful integration of windmills through energy systems. Developing accurate deterministic and probabilistic prediction methods for WP to facilitate efficient power system planning and operation has become increasingly important. A novel method that combines Variational Mode Decomposition (VMD) and adjusted LS-SVM for linear and stochastic interval prediction of a wind turbine [32]. Method applies VMD and adjusted LS-SVM to boost the precision and dependability of WP prediction, thereby enhancing power system management and decision making. VMD effectively manage the irregularity and instability inherent in WP series to combat the inherent variability. Additionally, an adjusted LS-SVM is utilized to create an accurate prediction model for WP that is resistant to anomalies and non-Gaussian error distributions [33].

In the global shift towards renewable energy production, wind energy has emerged as a significant and clean power source. However, the variable character of wind speed presents challenges for the generation of wind energy, as stochastic fluctuations can induce instability into the energy grid. As the demand for renewable energy sources continues to rise, investing in wind energy manufacturing carries enormous future potential. Therefore, accurate wind velocity predicting is crucial for the effective management of renewable energy development. Prediction wind speed to mitigate and reduce the uncertainties associated with wind power generation. Specifically, the work in [21], evaluates the application and performance of four distinct machine learning models for wind velocity prediction in the Las Vegas region of United States.

Wind power has emerged as a significant and rapidly growing energy source, fulfilling 16% of the EU's electricity demand [34]. However, the inherent volatility of wind power necessitates precise short-range prediction s for its successful integration of grid. Accurate hub height wind speed forecasts are essential for generating dependable wind power prediction s. Applying ensemble estimates obtained from numerous iterations of mathematical models for weather prediction is the current cutting-edge technique. Nevertheless, these ensemble forecasts frequently lack calibration and may exhibit biases, necessitating post-processing techniques to improve their predictive performance [35].

A case study is conducted using 100m wind speed estimates generated by the functional ensemble prediction method of the Hungarian Meteorological Department in order to evaluate its performance. Compared to three distinct ensemble model output statistics approaches and the raw ensemble forecasts, the prediction performance of the suggested strategy is evaluated. By addressing the calibration challenges associated with wind speed ensemble forecasts, it offers valuable insights into improving the reliability of wind power prediction s. Also a promising solution for enhancing the accuracy and usefulness of ensemble forecasts, this facilitates the integration of wind energy into energy production systems.

The difficulties in selecting suitable patterns of atmospheric circulation for the wind energy sector are examined in [17]. Initiative seeks to develop a user-friendly taxonomy of these patterns in the Euro-Atlantic region in order to mitigate the seasonal impact of climatic variation on the wind industry. To establish seasonal classifications with varying numbers of

clusters, the researchers apply K-means clustering to reanalysis data on sea level pressure. Effectiveness of the prediction system depends less on the number of clusters employed in the classification and more on the sea level pressure data's inherent precision, according to the findings. Analysis provides vital insights for mitigating the impact of atmospheric variability on wind energy generation and industry investment decisions [11].

Singh et al. presents a thorough analysis of methods for predicting and integrating wind Power. Fluctuating and irregular character of wind power poses difficulties for maintaining the secure and dependable functioning of power systems. Focus is on the most recent innovations in wind energy prediction techniques. Analysis covers a number of suggested and used strategies for prediction wind variations. Analysis also highlights several models such as RF regression, Long and Short Term Memory (LSTM), XGBoost, that are used to validate and simulate electricity power markets so that wind generation may be efficiently included [13]. Focus on the creation of novel models for prediction short-term wind power. A novel hybrid model is proposed that utilizes an optimization technique created especially for wind power prediction s [36].

In [37], growing importance of renewable energy sources is discussed, specifically wind, and solar energy, in the Polish power grid. For the conventional generators' schedules to be optimized and economic efficiency to be ensured, these sources' significant variability and poor ability to make precise prediction s of their future energy requirements using a variety of machine learning techniques. Extreme Gradient Boosting method turns out to be the top performance among the many machine learning approaches tested. Using two years of training data, the hourly output of wind energy in Poland for 2020 is predicted with a MAE of 26.7% and a RMSE of 4.5%.

Study reveals daily and seasonal fluctuations in the anticipated inaccuracy, with summer and daytime times showing larger MAE. These results emphasize how crucial it is to take temporal changes and specific time periods into account when estimating wind power. Study article provides insights into the development and evaluation of machine learning-based prediction models for day-ahead wind power prediction in Poland. Results contribute to a better understanding of wind power variability and decision-making processes for the efficient use of

conventional generators in the context of increasing renewable energy integration in the Polish electricity system [38].

Different setups and inputs for energy output is investigated in [15]. Traditional methods include comparing model sets to observations, which is not practical in offshore wind regions lacking reliable hub height data. As a result, the variability is modeled with wind velocities and conveying confidence in the absence of data. To overcome this problem, the WRF model is considered. Study examines techniques for attributing total ensemble variability to various ensemble elements [39].

To maximize the use of wind power, which is a growing renewable energy source, accurate wind speed forecasts are required. Numerous prediction models have been developed to improve the precision of wind speed and output prediction s. Nevertheless, the complexity of wind speed time series, which is characterized by nonlinearity, volatility, and intermittency, poses obstacles to precise forecasting. To resolve the aforementioned issues an efficient strategy is adopted to enhance the accuracy of the forecast. In addition, combination weights are calculated using partial least square regression which results in. MAE of 7.97% and 9.99% indicate that the proposed model is more accurate than competing methods [40].

Investigates an accurate wind speed and power prediction for ensuring stability, and extensive grid integration of wind power. A novel method for prediction is proposed that incorporates wavelet analysis, improved hybrid mode decomposition, and optimization techniques. Proposed method partitions wind speed data for hybrid mode decomposition and additional noise decomposition. After the wavelet analysis, an autoregressive moving average model is employed along with LSTM neural networks that are optimized with an improved particle swarm optimization method [41].

Difficulties associated with grid integration, load balancing, and energy trading brought on by the expansion of renewable energy facilities, as these issues need the creation of efficient prediction models. Recent research has highlighted the importance of utilizing spatio-temporal autocorrelation in plant data to enhance forecasts. However, the energy domain has not paid much attention to tensor models and approaches, despite their suitability for managing spatio-temporal data. To meet the requirements, a novel approach built on the Tucker tensor

decomposition is proposed [42]. Technique facilitates the derivation of a new feature space for the learning task by utilizing the benefits of tensor-based modeling. Researchers compare the efficacy of predictive clustering trees utilizing the new feature space versus the old feature space across three renewable energy datasets to evaluate the proposed strategy.

A novel technique for prediction wind energy system time series is proposed in [43]. Goal is to provide precise wind speed, produced power, and energy price forecasts in order to facilitate the most efficient operation, planning, administration, and marketing of wind energy systems. Proposed method employs a high-order neural network that has been taught the extended Kalman filter algorithm in real-time. In contrast to sophisticated hybrid methods or deep learning techniques, this method prioritizes implementation simplicity, computational simplicity, and real-time performance.

In [40], sparse machine learning is used to prediction wind power over the upcoming hour. Model builds a high-dimensional feature set that is solved with the sparse approach utilizing predicted values, real-time observations, and data from neighboring power plants. Proposed strategy is compared with a number of alternative approaches using actual wind power data from the NREL-118 test system [44]. Findings show that the proposed technique exceeds existing approaches in terms of prediction accuracy, outperforming broadcast values derived using meteorological and physical methods.

A hybrid prediction model [45] is used to boost the accuracy of wind speed prediction s, which is crucial for the effective integration of wind energy into the power grid. Model employs a decomposition strategy in order to split the input wind speed data. Relevant properties from each sub-series of data are then fed to deep neural networks as inputs. Wind speed data from the National Institute of Wind Energy is used to evaluate the effectiveness of the recommended method. Hybrid model is demonstrated to exceed established benchmark techniques in terms of prediction accuracy through thorough experimental assessments utilizing a variety of statistical indicators. Study advances wind speed prediction techniques and makes it easier to use wind energy as efficiently as possible in the power system.

Issues that arise from the growing integration of unconventional renewable sources, such wind energy, into power dispatch scheduling techniques is addressed in [46]. Traditional

methods of power dispatch scheduling presuppose controllability over energy sources, which is false for some renewable energy sources like wind energy that have intrinsic unpredictability. A learning-based system for creating ideal energy bands around wind energy estimates is presented in this study. Additionally, the suggested method may be expanded to combine several forecasts into a single prediction with a smaller band width and the same degree of confidence. Approach is developed and used using a real-world case study of the Uruguayan Electricity Market, which is well known for its significant penetration of renewable energies. By advancing the creation of energy bands, this research contributes to the effective management and integration of wind energy resources into the electrical grid.

Enhancing the wind energy power production systems' ability to anticipate short-term wind speed with greater precision is investigated in [41]. It is shown that the prediction models, which have poor accuracy since they ignore data pre-processing and depend only on one prediction algorithm. A novel approach to deal with these challenges is proposed that, combines five neural networks with powerful optimization algorithms and data pre-processing techniques. Proposed model is evaluated in four trials using wind speed data from China. Hybrid model outperforms other benchmark models in terms of accuracy and stability, as seen by lower levels of MAE and Std (Standard Deviation) performance indices.

2.2 Overview of Deep learning Methods

An extensive review of prediction wind energy using an Artificial Neural Network (ANN) is done in [47]. Results demonstrate the ANN is efficient in prediction wind energy. ANN provides prediction with accuracy and proper calibration with wind prediction instruments [48].

In [49], wind energy is predicted through time series using LSTM technique. By comparing this method with others better results are obtained. Yuan-Jia Ma et al employs a dual-step integrated machine learning (ML) model that combines the optimization with Feed-Forward Artificial Neural Network. Model utilizes the capabilities of these techniques to enhance the precision and effectiveness of wind energy forecasting. This involves two stages

first, estimate environmental factors (velocity, temperature, and so on). Second stage is to identify the optimal installation site for the wind form [50].

A. Khosravi et al. employs a variety of methodologies for prediction time series wind energy one approach is to combine a Multilayer Feed-Forward Neural Network with a Fuzzy Inference System. Another technique is Support Vector Regression (SVR) and an Adaptive Neuro-Fuzzy Inference System. These method groups are implemented to accurately predict wind speed patterns over time [51].

Yi Zangh et al. use ANN-based prediction model for wind energy prediction [52]. This is carried out on on hourly data. Method is based on time series statistics. Numerical Weather Prediction (NWP) and ANN are used to predict wind prediction. Wind speed and power are forecasted using real operating data. Prediction technique is built as a reference model. Findings indicate that the suggested grey combination model enhances prediction accuracy. In [52], statistical methods are combined with artificial intelligence techniques, including deep learning, to enhance energy forecasting. Nonlinear features and invariant datasets are utilized to improve accuracy. Findings provide a comprehensive analysis and discussion of various prediction techniques based on deep learning.

Lin Wang et al. apply Bidirectional Long Short-Term Memory Network (BiDLSTM) and wavelet transform are used to break down data for the purpose of wind energy prediction s. Experimental evidence indicates that the BiDLSTM method obtains superior prediction precision [53]. In [54], Empirical Mode Decomposition (EMD) model is utilized to determine in many phases the energy forecasting. In every strategy, the nonlinear wind speed is separated into three components. Development of the comparable EMD provided new insights into the data structures involving three years of wind speed data. These models' performance is measured by a combination of MAE and Root Mean Square Deviation **(**RMSD).

Wind energy prediction is made using a truncated model that incorporates area and wind data [55]. An adaptive wavelet neural network is used for examining wind speed, direction prediction s to wind energy forecasting. Since wind direction is fundamentally a circular variable, a modified type of wind direction variable is employed as an input for improved training and function approximation [32]. To integrate the LSTM network with an improved

Backpropagation neural network, a novel hybrid wind energy prediction model is proposed [56]. Using principles of deep learning and an improved Backpropagation algorithm, the model can accurately predict nonlinear wind energy.

Numerous methods, including Singular Spectrum Analysis along with complete ensemble empirical mode decomposition of adaptive noise, are employed to denoise and decompose the original wind speed data into various parts [40]. Procedure simplifies data characteristics and enhances the signal-to-noise ratio. Fuzzy Entropy is employed to assess the time difficulty of every element, and then the intrinsic mode function features are mixed again using the Spearman correlation method to create new subsequences. This reduces error accumulation and redundant computation. Proposed model integrates LSTM for predicting high-complexity subsequences with enhanced model for predicting low-complexity subsequences. Final estimates are then derived by combining the prediction s from the two models. Results of the experiment show that the proposed model outperform other models for prediction, exhibiting the least RMSE and MAE numbers. Moreover, a significant high correlation by Pearson appears between the predicted and actual wind speed values.

Due to the inherent unpredictability of wind energy, the rapid worldwide growth of wind energy over the past several decades has presented integration challenges. Therefore, precise real-time forecasts of outputs are necessary. Combining NWP and methods based on ML, specifically ANNs, to improve the accuracy of prediction wind power has emerged as a promising strategy. Two composite models integrating NWP and ANN are used study for prediction wind power in complex terrains [57]. It is shown that proposed models produce accurate wind farm power forecasts, particularly in highly complex terrains with MAE of 8.76% and a Root Mean Squared Error of 13.03%. These results demonstrate the viability of the suggested wind power prediction models for complex terrains.

As a clean and renewable energy source, wind power has seen unprecedented growth. Accurate wind power/velocity interval prediction is necessary for the efficient dispatch of wind energy. Extensively utilized Lower Upper Bound Estimation model for interval prediction is a crucial method for energy forecast. Novel Huber loss function is discovered to be more efficient than conventional loss functions as the research suggests and analyzes various loss function types. Enhanced the model is then built. Showing potential for improving the effectiveness and

accuracy of wind energy [58]. Relative efficiency of AI approaches for wind power prediction in Portugal is compared in. Include ANNs and Radial basis function network with various learning methods. Results demonstrate the superior performance of ANN in this specific circumstance. Findings contribute to an increase in the precision of wind power forecasts, allowing for enhanced wind energy resource planning and utilization [59].

Lang et al. focus on improving wind power forecasts for the stability and security of grid operation. Due to the limitations of conventional wind power point prediction s, a new two-stage short-term wind power interval prediction method was developed which combines the minimum gated memory network with an enhanced interval width adaptive adjustment algorithm. A network-based point model is first established for representing the subsequences of wind power data. Enhanced interval width adaptive adjustment technique, which changes the prediction interval labels, is then used to suggest an interval model in order to derive the final prediction intervals. Proposed model enhances the precision and dependability of wind power interval forecasts by integrating the network with a more effective adaptive interval width adjustment technique. These findings have implications for enhancing the integration of wind power into grid systems and promoting the efficient use of renewable energy sources [60].

The importance of ramp event and short-term wind power projections for effective risk management and grid operation in smart grids is highlighted in [61]. A hybrid strategy based on a semi-supervised generative adversarial network has been proposed to address these prediction issues. Model divides the original time series of wind energy into smaller time series with varied frequencies. It uses labelled learning and semi-supervised regression to identify the nonlinear and dynamic characteristics of each. Data clustering properties of wind power outputs are captured using the GAN generative model by generating unlabeled virtual samples. Additionally, a self-tuning prediction approach with a multi-label classifier is applied to help with the prediction of wind power peak occasions. Results demonstrate the superior performance of the suggested approach in the context of wind power prediction and provide valuable information to improve grid operation and risk management.

The importance of accurate wind forecasts for the reliability of wind energy integration in power systems is discussed in [62]. Changes in turbine power output can disrupt the equilibrium between energy demand and supply, making accurate prediction essential for

efficient power system planning and continuous supply. This is accomplished by combining the strengths of LSTM and genetic algorithm for short-term wind power prediction in a framework called genetic long-term memory. It is shown that model's performance is better than support vector regressor forecasts, and other documented methodologies. It has practical implications for augmenting the dependability and effectiveness of wind power integration into power systems, thereby facilitating the efficient planning and supply of energy.

Wind energy is highlighted as a significant source of renewable energy for electricity generation and other purposes. For wind power generation and other applications, it is necessary to accurately prediction the wind speed due to its unpredictability. ANN and multiple linear regression (MLR) are evaluated for predicting wind speed in the central region of Chhattisgarh, India in [47]. It take into account relative humidity, wind speed, ambient temperature, ambient pressure, and discernible water. A Multilayer Perceptron model is trained with a 5-20-1 architecture. Model obtains the lowest RMSE and mean relative error values of 0.4558 and 0.15, respectively, with a correlation coefficient of 0.90162. Wind speed is the dependent variable, and the MLR method employs the same parameters, yielding a correlation coefficient of 0.77852.

The difficulties posed by the erratic and volatile nature of large-scale wind energy's incorporation into contemporary power networks. Wind power prediction is crucial for resolving these issues because it provides comprehensive information about potential future uncertainties in wind energy. In the context of wind power forecasting, an in-depth and timely assessment of meta-heuristic methods is done in [63]. Three layers that make up the framework are the auxiliary layer, prediction base layer, and core layer is used. Several error evaluation metrics, such as deterministic, indeterminate, and testing approaches, are discussed in the context of wind power forecasting. In a quantitative investigation, the advantages, disadvantages, prediction accuracy, and computational costs of various algorithms are highlighted. In addition, the report emphasizes current trends and outstanding research concerns, which aids the reader in understanding each wind power prediction technique.

The challenges of wind power prediction and the existence of anomalies in actual wind power statistics due to ambiguous causes. In the presence of anomalies and non-Gaussian error distributions, conventional prediction techniques based on mean square error (MSE) loss are

insufficient [24]. By combining entropy with LSTM neural networks, it is suggested that reliable short-term wind power hybrid prediction can be achieved. Method is named as improved variation mode decomposition and Sample Entropy approach. Proposed approach improves prediction accuracy by considering outliers and non-Gaussian error distributions into consideration.

Regarding the significant nonlinearity and non-stationary exhibited by wind speed due to the effect of the atmospheric boundary layer, the need for precise and stable wind speed forecasts for the safety of power infrastructures must be emphasized. In order to increase prediction accuracy, this a novel hybrid prediction system that incorporates efficient data decomposition techniques, recurrent neural network is used for prediction [24]. Results demonstrate that the proposed hybrid system is superior to other single models and conventional methods, producing extremely precise wind speed forecasts. Results demonstrate the predictive accuracy advantage of the proposed system.

Featuring a special emphasis on the Pir Panjal Range in the Himalayan region, the information is used to predict wind energy, which is crucial for maintaining the consistency of power production and coordinating the future use of wind energy in [64]. ANN is used on a 30-day dataset of wind speed, temperature, and air density for training and subsequent wind energy prediction. Utilizing ANN for wind energy forecasting, this study contributes to the field of wind energy prediction in the Himalayas. Validating th0065 prediction model entails evaluating the results and comparing them to actual data. Study provides insightful data on the future potential for wind energy production in mountainous regions, allowing for more precise planning and increasing electrical system dependability.

The increasing use of wind energy in power networks has increased the significance of wind speed prediction s. Due to variations in wind speed time series, the prediction techniques demonstrated low accuracy. An optimal sub-model is used based on modified multi-objective optimization method is used for point prediction, interval prediction is based on distribution fitting, and a system assessment is provided as a solution to this problem. Experimental results demonstrate the efficiency of the proposed method, with absolute percentage error values for one-, two-, and three-step point prediction s for Site 1 and Site 2 of 2.9220, 3.1696, and 4.8355, respectively [65].

A novel Markov prediction model that integrates wind acceleration data to improve prediction precision is proposed in [66]. Proposed method encodes the wind speed sequence into a discrete state sequence. From this discrete state sequence, the Transition Probability Matrix, which governs state transitions in the Markov chain, is computed. Model is essential for prediction future conditions. Projected state sequences are translated into wind speed, and predictive distributions are used to quantify the prediction uncertainty. Suggested approach has benefits in terms of increased prediction accuracy and adaptability in adding other input to the model. Case studies are used to confirm the method's efficacy, and it is compared to other approaches. Overall, the suggested Markov prediction model shows encouraging results and helps the wind sector enhance wind speed prediction .

Application of big data and deep learning techniques to energy prediction for improved grid planning, operation, and management. To enhance renewable energy integration and grid modernization, the study in [1] analyzes AI and ML approaches, emphasizing their benefits and drawbacks. Focus is therefore focused on the capacity of DL algorithms to manage large data sets and automatically extract nonlinear properties. Combining Markov error correction, Fuzzy neural network, and improved genetic algorithm optimized complementary empirical mode decomposition is proposed as a novel combination model in [67]. Prediction outcomes are then merged, and Markov error correction is implemented. Proposed model yields reduced values for MAE, and RMSE i.e. 15.59%, and 17.95%, respectively. Proposed technique significantly reduces MAE, RMSE, and values, demonstrating its superior prediction accuracy. By providing more precise ultrashort-term wind energy projections, the findings of this study demonstrate the potential of the proposed technique for the expansion and use of wind power.

Development of ANN models for prediction wind energy generation is used for a wind farm in Sri Lanka [68]. Performance metrics such as MSE and RMSE is calculated as with R greater than 0.91, RMSE equal to 0.22, the results demonstrate that working of ANN model. Existence of numerous prediction methodologies as well as the nonlinear and time-varying link between wind and produced electricity. Idea of merging numerous forecasts for the same hour and prediction horizon is explored in order to increase prediction accuracy in [69]. Markov chain models for combination and multivariate dimension reduction approaches are used. It is shown that it efficiently enhances wind power prediction s.

Prediction with precision is focused in [70] by using LSTM, and Convolutional Neural Network models. Experimentation is conducted to create a hybrid model that combines the benefits of the individual models. SCADA system data is used for training and validation purposes. Based on evaluation metrics it is shown that the hybrid model gives MAE values of 0.1365, RMSE values of 0.0974. Hybrid model outperforms conventional wind power prediction models in terms of prediction precision.

2.3 Summary

Literature review demonstrates the effective implementation of use of the ML method for wind energy output prediction. ML models have the potential to prediction wind energy with accuracy, which is crucial for effective energy planning, grid management, and the seamless integration of renewable energy. Such models can be used by energy planners, grid administrators, and regulators to inform decisions regarding the integration of renewable energy sources and overall energy management. Outcomes paves the way for ongoing improvements in wind energy prediction and the greater acceptance of sustainable energy sources by academics, practitioners, and policymakers in the field of renewable energy

.

CHAPTER 3

METHODOLOGY

This chapter presents the methodology used in this thesis. This includes datasets, ML methods, performance metrics and the proposed method. Figure 3-1 shows the block diagram of the proposed method. Detail about each block is given the subsequent sections

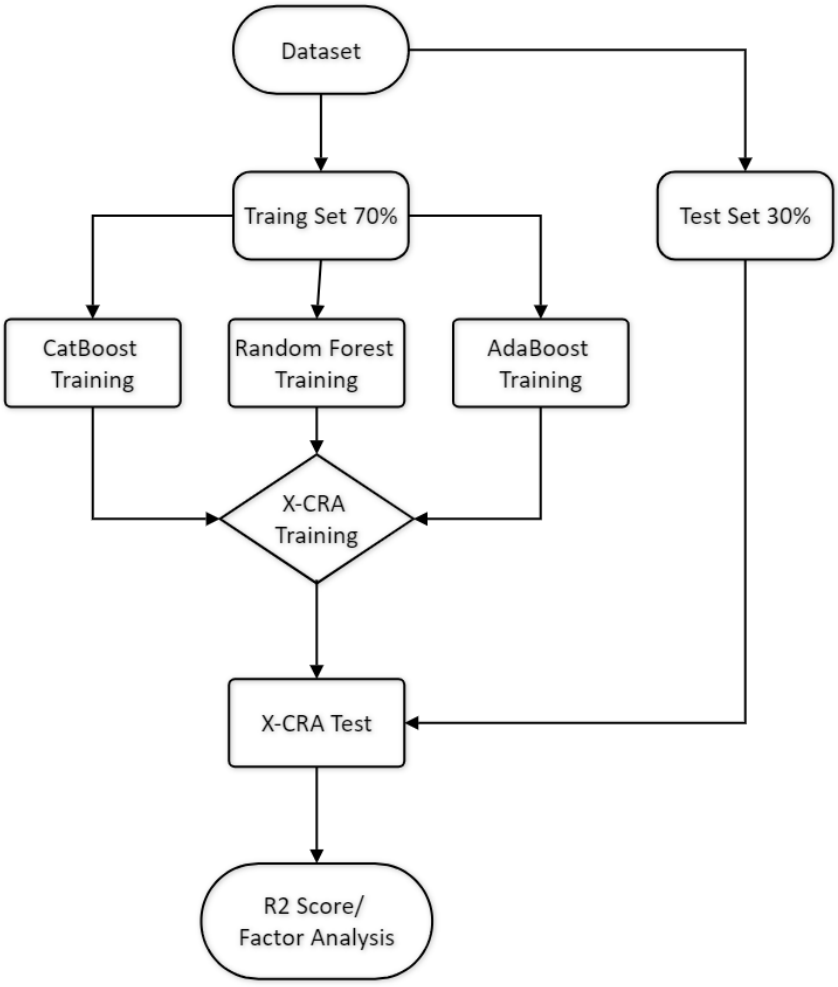

Figure 3.1: Block diagram of the proposed X-CRA method

3.1 Proposed Method (X-CRA)

Proposed methods combines the strength of XGboost, CatBoost, Random Forest AdaBoost. The proposed method is names as (X-CRA). CatBoost, Random Forest and AdaBoost methods are first independently trained one by one to predict the energy output. Their prediction is combined through the stacking step of the ensemble learning using the XGBoost method. It means that their prediction are fed into XGBoost which combine them to produce a single output i.e prediction of wind energy output. X-CRA method is denoted as CatBoost + Random forest + Adaboost. Additionally, Random Forest + AdaBoost, CatBoost + AdaBoost, and CatBoost + Random Forest are also evaluated and compared with X-CRA.

3.2 Datasets

Three publicly available datasets are used in this thesis which are listed in Table 3-1. Detail about each dataset is present in the subsequent sections.

Table 3.1: List of used datasets

Sr. No.	**Dataset**	**Year**
1.	Wind power curve modeling Dataset [71]	2021
2.	A Fine Windy Day Dataset [72]	2021
3.	KDD Dataset [73]	2022

3.2.1 Wind Power Curve Modeling Dataset

Wind power curve modeling dataset [71] is a comprehensive and a valuable resource for analyzing features and factors influencing wind turbine output and advancing wind energy generation research. Dataset encompasses sensory data collected through various sensors to predict wind energy production.

Data collection process involved the utilization of wind turbines equipped with sensors and monitoring systems. These devices capture critical parameters and environmental variables, allowing for a comprehensive assessment of wind energy generation. Dataset's primary objective is to model the relationship between the power output of wind turbines and the corresponding wind speed. Dataset contains two features (factors), each playing a distinctive role in shaping wind energy production as shown in Table 3-2.

Table 3.2: Wind Power Curve Modeling Dataset

Sr. No.	Feature name	Description
1.	Wind Speed	This parameter is a pivotal determinant of wind turbine output. Wind speed affects the rotational speed of the turbine's blades, directly impacting energy generation. As wind speed increases, the turbine's power output generally rises due to increased mechanical energy.
2.	Wind Direction	Wind direction influences the efficiency of energy capture by determining the angle of attack of the turbine blades. Optimal alignment with wind direction enhances energy extraction.

Each of these features holds critical importance in wind energy generation. Wind speed and wind direction influence the amount of kinetic energy available for conversion into electricity.

3.2.2 A Fine Windy Day Data

A Fine Windy Day Data [72] is a comprehensive and a valuable resource for analyzing features and factors influencing wind turbine output and advancing wind energy generation research. Dataset encompasses sensory data collected through various sensors to predict wind energy production. It includes features such as Table 3-3.

Table 3.3: A Fine Windy Day Dataset

Sr. No.	Feature name	Description
1.	Wind Speed	Wind speed is a critical factor directly affecting wind energy generation. Higher wind speeds generally result in greater energy production.
2.	Atmospheric Temperature	Atmospheric temperature can influence air density, which impacts turbine performance. Cooler temperatures may result in denser air and potentially higher energy output.
3.	Shaft Temperature	Shaft temperature is indicative of the turbine's mechanical condition. Overheating may affect efficiency and require maintenance.
4.	Blades Angle	The angle of turbine blades affects the amount of wind captured. Optimal angles optimize energy capture.
5.	Gearbox Temperature	Gearbox temperature is crucial for smooth turbine operation. High temperatures could impact efficiency and longevity.
6.	Engine Temperature	Engine temperature can affect overall turbine performance and maintenance requirements.
7.	Motor Torque	Motor torque is directly related to energy conversion. Higher torque suggests more efficient energy generation.
8.	Generator Temperature	Generator temperature affects energy conversion. Cooler temperatures can enhance efficiency.
9.	Atmospheric Pressure	Atmospheric pressure influences air density and thus turbine performance.
10.	Area Temperature	Area temperature may impact overall turbine efficiency and performance.
11.	Windmill Body Temperature	The temperature of the windmill body is indicative of its operational state and potential maintenance needs.
12.	Wind Direction	Wind direction affects turbine orientation for optimal energy capture.
13.	Resistance	Resistance is relevant for electrical performance. Deviations may indicate issues in the system.
14.	Rotor Torque	Rotor torque affects energy conversion and overall turbine performance.
15.	Turbine Status	This categorical feature may indicate various operational states of the turbine, which can influence energy generation.
16.	Cloud Level	Cloud cover may affect incoming wind patterns and energy generation.
17.	Blade Length	Blade length impacts the swept area and energy capture.
18.	Blade Breadth	Blade breadth affects energy capture and turbine efficiency.
19.	Windmill Height	Windmill height can impact exposure to different wind speeds and directions.
20.	Windmill Generated Power	Target variable Represents the windmill's actual generated power and is the key metric for wind energy analysis.

Each of these features plays a crucial role in wind energy generation and turbine performance. By analyzing and understanding their relationships, an optimized turbine settings can be achieved to maximize energy output. Additionally, employing regression techniques can help in uncover the most significant factors driving energy generation, aiding in the advancement of renewable energy technologies [72].

3.2.3 KDD Dataset

KDD Dataset [73] is a latest, comprehensive and valuable resource for analyzing features and factors influencing wind turbine output and advancing wind energy generation research. Dataset encompasses sensory data collected through various sensors to predict wind energy production. It includes features such as Table 3-4.

Table 3.4: KDD Dataset

Sr. No.	Feature name	Description
1.	Wind Speed	It directly influences the kinetic energy available in the moving air, which is harnessed by the turbine's blades to generate power. Higher wind speeds typically result in increased power output.
2.	Wind Direction	The angle between the wind direction and the position of the turbine nacelle is crucial for efficient energy capture. Accurate adjustment of the nacelle direction based on wind angle enhances energy conversion efficiency.
3.	Temperature	The temperature of the surrounding environment influences the air density, which has a significant impact on turbine performance.
4.	Temperature of Inside Nacelle	The temperature inside the turbine nacelle affects various components and systems. Maintaining optimal nacelle temperature is crucial for sustained energy production.
5.	yaw angle	The yaw angle is the angle between the rotation direction of the nacelle and the relative wind vector.
6.	Pitch Angle	The pitch angle refers to the angle at which the turbine blades are set relative to the wind direction. Adjusting the pitch angle allows control over energy extraction.

3.3 Preprocessing of Dataset

All the datasets described above require the preprocessing of the datasets to prepare them for ML model.

3.3.1 Data Cleaning

During the data cleansing process, it is critical to understand the data type of the features and their probable range of values. Without this information, distinguishing between acceptable and undesirable values becomes difficult. Missing values, commonly known as NaN and Null, are removed during the data cleaning stage to guarantee that the dataset used for ML is full and contains numeric values [74].

3.3.2 Feature Engineering

Feature engineering is a method of transforming, and constructing features convert them into numeric values such as label encoding [75].

3.3.3 Feature Scaling

Scaling of features is a preprocessing step. Standardized scaling techniques changes the data to have the mean of 0 and a standard deviation of 1, This help in early convergence of the training ML models and also result in good performance [76].

3.3.4 Train-Test Split

Train test split of the above dataset is 70% and 30%, respectively. 70% data is used for the training of the ML models and the remaining 30% dataset for testing.

3.4 Machine Learning Algorithms

ML methods used are explained in the following sections. These ML methods are trained as a regressor to predict the output energy, which is of continuous data type.

3.4.1 Decision Tree

A decision tree [78] is a ML approach that uses a succession of if-then-else choices to determine the label or value for a given observation. Usage of DT has the advantage of being simple to read and may be used to model non-linear connections. The parameters used for DT are criterion: gini, splitter: best, min_samples_split:2, min_samples_leaf:1.

3.4.2 Random Forest

Random Forest [79] is a well-known ML method. It is built by growing multiple decision trees. Each decision tree gives a prediction which is aggregated in an ensemble way to find the best results. Fact that random forest splits features into random subspaces to capture non linearity in the data and which lead to good generalization error. The Parameters used used for RF are n_estimators:100, criterion:gini, min_samples_split:2, min_samples_leaf:1, min_weight_fraction_leaf:0.0

3.4.3 Extra Tree

Extra Tree [80] is a powerful ML method. Extra Tree constructs an ensemble of decision trees to make prediction s. However, what feature sets or value it uses for splitting a node is totally random and not by using any impurity calculation. This is to encapsulate the randomness in the data and make low generalization error on unseen data. Unlike traditional decision trees that choose the best split based on a certain criterion, Extra Tree selects splits at random. Extra Tree is particularly well-suited for handling noisy and complex datasets, as it embraces the variance in the data to create a more robust predictive model. Its versatility and ability to capture non-linear relationships between output and input variables. The Parameters used for Extra Tree are n_estimators:100, criterion:gini, max_depth:None, min_samples_split:2, min_samples_leaf:1, max_features:sqrt.

3.4.4 Gradient Boosting Regression

Gradient Boosting Regression [81] is a powerful ML method for nonlinear data and regression tasks. Method belonging to the ensemble learning category, combines multiple weak learners, primarily decision trees, and sequentially form a predictive model. By compiling the residual errors of preceding models, Gradient Boosting progressively refines its predictive power, concentrating on instances that were inaccurately forecasted. Iterative enhancement process effectively minimizes the loss function and accurate regression model capable of understanding complex data patterns. The parameters Used for Gradient Boosting Regression are loss:squared_error, learning_rate:0.1, n_estimators:100, subsample:1.0, criterion:friedman_mse, min_samples_split:2, min_samples_leaf:1

3.4.5 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a strong and efficient gradient boosting algorithm implementation [82]. XGBoost is well-known for its speed and scalability. It entails adding weak models to the ensemble repeatedly, with each successive model aiming to rectify

the mistakes caused by the prior models. XGBoost steadily enhances the ensemble's overall prediction performance by learning from the remaining results of prior models. XGBoost employs decision trees as base learners to create a more effective ensemble, with each tree contributing to the final prediction depending on its own strength. The Parameters used in XGBoost are oss:log_loss:, learning_rate:0.1, n_estimators:100, subsample:1.0, criterion:friedman_mse, min_samples_split:2, min_samples_leaf:1

3.4.6 **Light Gradient Boosting Regression**

Light Gradient Boosting Model (LightGBM) is a cutting-edge machine learning model that has gained significant attention [83]. As a variant of gradient boosting, LightGBM is implemented to optimize both efficiency and performance. It achieves this by employing a novel approach to constructing decision trees, using a histogram-based algorithm that reduces memory usage and speeds up training. LightGBM's distinctive feature lies in its ability to find the optimal split points for continuous features, leading to more accurate and efficient tree building. LightGBM emerges as a compelling choice for tackling complex problems across various domains. The parameters used in LightGBM are boosting_type:gbdt, num_leaves:31, max_depth:-1, learning_rate:0.1, n_estimators:100, subsample_for_bin:200000

3.4.7 **Categorical Boosting**

CatBoost [84] is a cutting-edge ML method. CatBoost is uniquely designed to handle categorical features seamlessly, which often pose challenges in traditional gradient boosting methods. By employing an innovative approach that incorporates ordered boosting and decision trees, CatBoost effectively captures intricate relationships within categorical data. CatBoost’s inherent capability to handle categorical features, along with its efficient handling of missing values, makes it effective for the regression tasks. The Parameters used in CaBoost are loss_function:RMSE, and all other parameters are set as default.

3.4.8 Adaboost

Adaboost [85] is a short form of Adaptive Boosting, is a prominent and influential machine learning algorithm renowned for its capability to enhance the performance of weak learners and produce a strong, accurate predictive model. Adaboost operates by iteratively adjusting the weights assigned to training instances, focusing on those that are misclassified in each round. Iterative process effectively prioritizes challenging examples and compels subsequent weak learners to improve their accuracy on these instances. By combining multiple weak learners, typically decision trees with limited depth, Adaboost creates a robust ensemble model that excels in handling complex datasets and capturing intricate patterns. Its adaptability to various domains and flexibility in accommodating different base learners make Adaboost a widely utilized algorithm in classification and regression tasks. Despite its sensitivity to noisy data, Adaboost's effectiveness in boosting overall predictive performance has solidified its place as a fundamental technique in the machine learning toolbox. The parameters used in Adaboost are n_estimators:50, learning_rate:1.0, algorithm:SAMME.R.

3.5 Performance Metrics

Performance analysis is carried out by using following metrics.

3.5.1 Coefficient of Determination (R^2)

R^2 score is also called the coefficient of determination. It provides valuable insights into the proportion of the variance in the wind turbine output that can be explained by the regression model [86]. It is used to compare the regressive value with actual l value. R^2 score has max minimum score of 1 and min of 0. Maximum value of 1 is indication of both prediction and actual being perfectly correlated and 0 as vice versa.

To compute the R^2 score, the first step involves fitting the regression model to the dataset containing the factors that potentially influence the wind turbine output. Once the model

is trained, it predicts the turbine outputs based on the given factors. R^2 score is then calculated by comparing the variation in the predicted turbine outputs to the actual turbine outputs[76]. Mathematically, the R^2 score is expressed as the ratio of the explained variance to the total variance.

$$R^2 = 1 - \frac{RSS}{TSS} \quad (3.1)$$

R^2= Coefficient of Determination

RSS= Sum of Squares of Residuals

TSS= Total Sum of Squares

R^2 score ranges between 0 and 1, where a score of 1 indicates that the model perfectly explains the variance in the data, while a score of 0 suggests that the model does not provide any meaningful explanation.

3.5.2 Root Mean Square Error

RMSE is a performance metric that aids in comprehending the predictive accuracy of the model [87]. RMSE measures the average magnitude of the residuals, which are the differences between the actual wind turbine output values and the corresponding predicted values provided by the regression model. Metric quantifies the model's ability to accurately estimate the turbine output.

Mathematically, the RMSE is calculated using the following formula [88].

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (3.2)$$

RMSE score provides a meaningful interpretation of the model's predictive accuracy. A lower RMSE indicates that the model's prediction s are closer to the actual turbine output values, signifying higher precision. Conversely, a higher RMSE suggests that the model's prediction s deviate further from the actual values, indicating reduced accuracy[89].

In the context of identifying the most influential factor for maximizing wind turbine output, RMSE plays a vital role. By calculating and comparing RMSE values for different factors, we can discern which factor contributes the least to prediction errors. Factors that result in lower RMSE scores imply a stronger correlation with wind turbine output and better predictive performance. Therefore, the factor associated with the lowest RMSE value is likely to be the most influential in driving turbine output[90].

3.5.3 **Mean Absolut Error**

In the context of regression analysis on wind turbine output, the MAE is a vital performance metric that offers valuable insights into the predictive accuracy of were model [91]. MAE quantifies the average magnitude of the absolute differences between the actual wind turbine output values and the corresponding predicted values provided by egression model. 3.3

$$\mathrm{MAE} = \frac{\sum_{i-1}^{n} |y_i - \hat{y}_1|}{n} \qquad (3.3)$$

MAE score offers a straightforward and intuitive interpretation of the model's predictive accuracy. A lower MAE indicates that the model's prediction s are closer to the actual turbine output values, reflecting higher precision. Conversely, a higher MAE suggests that the model's prediction s deviate further from the actual values, indicating reduced accuracy. In the context of identifying the most influential factor for maximizing wind turbine output, MAE plays a significant role. By calculating and comparing MAE values for different factors, we can determine which factor contributes the least to prediction errors. Factors associated with lower MAE scores demonstrate a stronger correlation with wind turbine output and better predictive performance. Therefore, the factor associated with the lowest MAE value is likely to be the most influential in driving turbine output [92].

CHAPTER 4

EXPERIMENTAL CONFIGURATION, FINDINGS, AND ANALYSIS

Chapter presents the experimental results. In total three wind energy datasets are used. On each dataset different regression techniques have been applied for wind energy prediction. In the following sections results achieved on each dataset are presented

4.1 Comparison of Machine Learning methods

In this section comparison of different ML methods for wind energy prediction is presented on three different datasets.

4.1.1 Results Obtained on Wind Power Curve Modelling Dataset

Results of ML methods using the Dataset name Wind Power Cure Modelling are shown in Table 4-1.

Table 4.1: Results achieved on wind power curve modeling dataset

Method	Train set	Test Set		
	R^2	R^2	MAE	RMSE
CatBoost	0.993	0.992	0.075	0.109
XGBoost	0.993	0.991	0.085	0.116
Decision Tree	1.000	0.986	0.098	0.149
Random Forest	0.997	0.991	0.084	0.117
GBR	0.992	0.991	0.085	0.116
LGBM	0.997	0.991	0.084	0.117
ExtraTree	1.000	0.991	0.079	0.119
AdaBoost	0.984	0.984	0.123	0.157

- **CatBoost**

CatBoost, a robust method known for its effectiveness with categorical features, is applied to the Wind Power Curve Modelling dataset. Impressively, it achieved an R-squared (R2) score of 0.993 during training and 0.992 on the test set, indicating its ability to forecast. Moreover, CatBoost displayed remarkable precision, yielding MAE values of 0.075 and RMSE values of 0.109 for testing, respectively. These outcomes demonstrates CatBoost's ability in capturing intricate wind power dynamics, positioning it as a potent tool for enhancing renewable energy modeling accuracy.

- **XGBoost**

Employing the Wind Power Curve Modelling dataset, the XGBoost method demonstrated its effectiveness. Notably, it attained a high R-squared (R^2) score of 0.993 during training and 0.991 on the test data, underscoring its ability to capture variance and generalize effectively. Additionally, XGBoost showcased precise prediction s with MAE values of 0.085 and RMSE values of 0.116 for testing, reinforcing its capacity to model intricate wind power dynamics accurately.

- **Decision Tree**

DT method exhibited strong performance. Notably, it achieves a perfect R-squared (R^2) score of 1 during training, showcasing its ability to precisely capture training data patterns. During testing, the model maintained its effectiveness with an R^2 score of 0.986, indicating robust generalization. DT method demonstrated a MAE of 0.098 on the testing set and RMSE of 0.149 on the test set, underscoring its potential to accurately model and predict wind power dynamics.

- **Random Forest**

RF machine learning method emerges as a robust method for wind power prediction. During training, Random Forest achieves an impressive R-squared (R^2) score of 0.997, showcasing its adeptness in capturing intricate data patterns. As the test phase unfolds, the model maintains a substantial R^2 score of 0.991, signifying its capability to generalize and predict wind power effectively. Notably, Random Forest demonstrates accuracy in prediction, evidenced by MAE values of 0.084 and Root Mean Square Error values of 0.117 during testing. These results underline Random Forest's potential to model wind power dynamics with precision.

- **GBR**

Gradient Boosting Regressor (GBR) method demonstrates good capabilities. During the training phase, GBR achieved a high R-squared (R^2) score of 0.992, indicating its adeptness in capturing data patterns. Its efficacy carried over to the test phase, maintaining a strong R^2 score of 0.991, showcasing robust generalization. GBR method also showcased accurate prediction s, as evident from its MAE values of 0.085 on training data and 0.116 on the test data. These results highlight GBR's potential to accurately model and predict wind power dynamics, contributing to the field's advancement.

- **LGBM**

LightGBM (LGBM) method demonstrates good performance. During training, LGBM achieved an impressive R-squared (R^2) score of 0.997, showcasing its remarkable ability to capture intricate data patterns. Its efficacy extended to testing, maintaining a strong R^2 score of 0.991, indicative of robust generalization. Notably, LGBM showcased precise prediction s with MAE values of 0.084 and RMSE values of 0.117 for testing data. These results underscore LGBM's capability to accurately model and predict wind power dynamics, contributing significantly to the field's advancement.

- **Extra Tree**

Extra Tree method exhibits good performance. During training, Extra Tree achieved a perfect R-squared (R^2) score of 1, signifying its exceptional ability to precisely capture training data nuances. Efficacy extended to the test phase, with a robust R^2 score of 0.991, indicating strong generalization capabilities. Additionally, Extra Tree demonstrated accurate prediction s, evident from its low MAE values of 0.079 and RMSE of 0.119 for testing. These outcomes underline Extra Tree's potential to adeptly model and predict wind power dynamics, showcasing its valuable contribution to the field's advancement.

- **AdaBoost**

AdaBoost method demonstrates relative low performance. During training, AdaBoost achieved a notable R-squared (R^2) score of 0.984, indicative of its ability to capture training data patterns effectively. Performance extended seamlessly to testing, maintaining a parallel R^2 score of 0.984, reflecting consistent generalization capabilities. AdaBoost showcased its predictive accuracy with MAE values of 0.123 and RMSE values of 0.157 on the test data. These results highlight AdaBoost's potential to model and predict wind power dynamics, contributing significantly to the field's advancement.

4.1.2 Feature Importance Based on Wind Power Curve Modelling Dataset

Examining feature importance across various regressors on the Power Curve Modelling dataset provides invaluable insights into the determinants of wind turbine output. Synthesis of average overall regressor importance offers a comprehensive understanding of these dependencies, shedding light on the key factors shaping wind turbine performance. Within this context, CatBoost, Random Forest, AdaBoost, XGBoost, Decision Tree, Gradient Boosting Regressor (GBR), LightGBM (LGBM), ExtraTree, and Ensemble methods collectively reveal the influential contributors to wind turbine output. Consistently across these regressors,

attributes such as wind speed, wind direction, temperature, and air pressure emerge as primary drivers.

Feature importance by different regression on power curve modelling datasets is shown in Figure 4.1. Average overall regression are also shown.

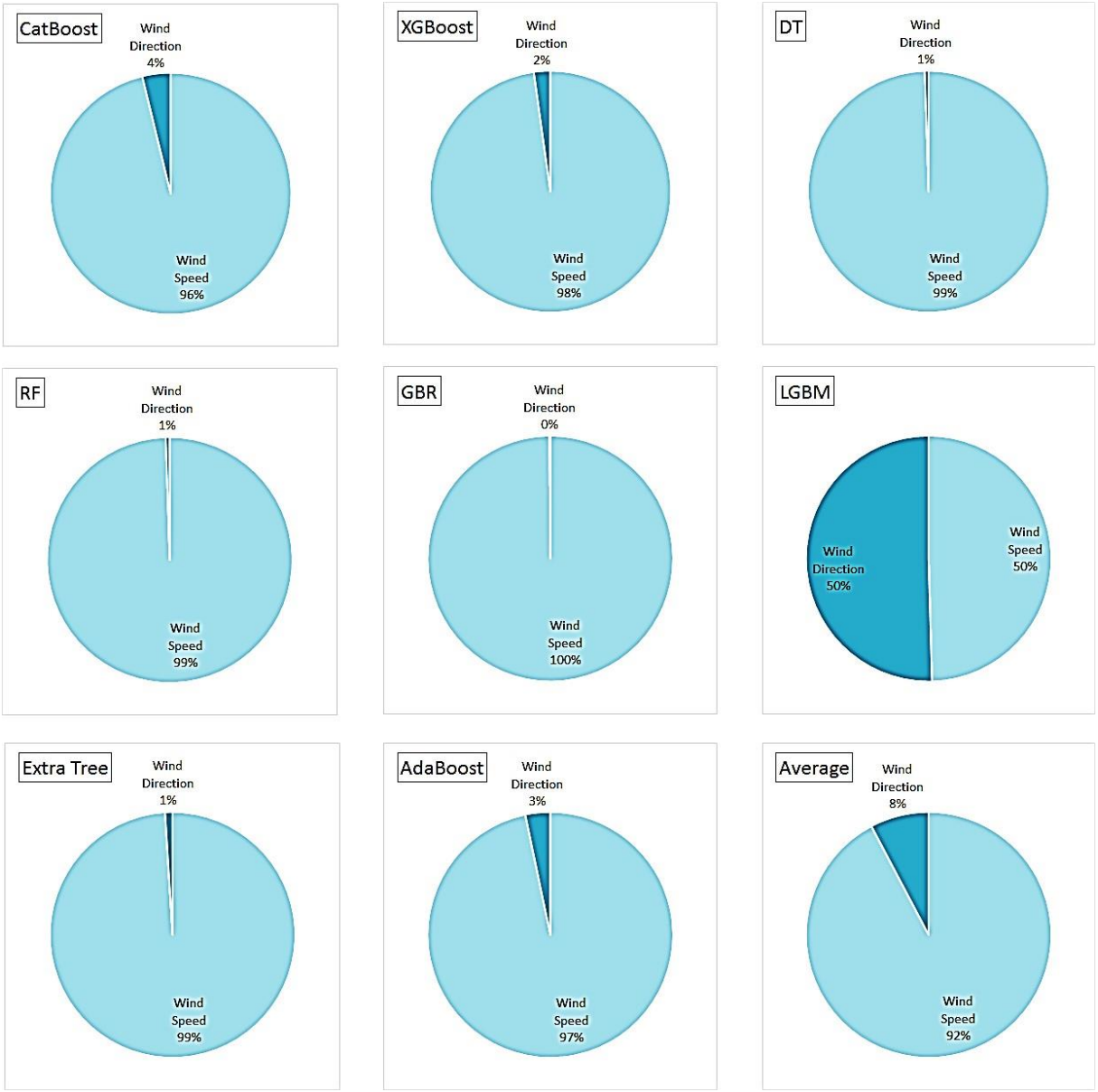

Figure 4.1: Feature importance of Wind Curve Modeling Dataset

Wind speed, acting as the central force behind turbine rotation, consistently garners significant importance. Wind direction, guiding optimal turbine orientation, follows closely in prominence. Moreover, environmental conditions like temperature and air pressure, impacting

air density and turbine efficiency, hold notable sway. In considering the ensemble of regression and calculating the average overall importance, a unanimous consensus forms around the centrality of wind speed, wind direction, temperature, and air pressure. Convergence amplifies the significance of these core variables in influencing wind turbine output. Analysis of feature importance by diverse regression on the Power Curve Modelling dataset underscores wind speed, wind direction, temperature, and air pressure as critical determinants of wind turbine output. Unified findings across various models reaffirm the importance of these factors, providing valuable insights for optimizing wind energy generation. By comprehending these dependencies, stakeholders can make informed decisions to enhance turbine efficiency, contributing to the advancement of sustainable energy solutions.

4.1.3 Results Obtained on a Fine Windy Day Dataset

Results of ML methods using the Dataset name A Fine Windy Day are shown in Table 4-2.

Table 4.2: Results achieved on fine windy day dataset

Method	Train Set	Test Set		
	R^2	R^2	MAE	RMSE
CatBoost	0.993	0.964	0.304	0.518
XGBoost	0.956	0.948	0.400	0.627
Decision Tree	1.000	0.924	0.423	0.753
Random Forest	0.995	0.962	0.300	0.535
GBR	0.954	0.945	0.412	0.642
LGBM	0.976	0.962	0.329	0.535
ExtraTree	1.000	0.962	0.334	0.534
AdaBoost	0.766	0.768	1.030	1.319

- **CatBoost**

A fine windy day dataset, the CatBoost machine learning method is as a powerful method for wind power prediction. During the training phase, CatBoost achieves an impressive R-squared (R^2) score of 0.993, indicating its adeptness in capturing underlying patterns within the data. As the test phase unfolds, CatBoost maintains a robust R^2 score of 0.964, showcasing its ability to generalize effectively and provide reliable prediction s. Precision of CatBoost's prediction s is evident through its MAE values, registering at 0.304 and RMSE values of 0.518 during testing. These results underscore CatBoost's proficiency in modeling wind power dynamics accurately, even on the finer nuances of a windy day.

- **XGBoost**

The XGBoost machine learning method emerges as a robust contender for wind power prediction. During the training phase, XGBoost achieves a commendable R-squared (R^2) score of 0.956, signifying its capability to capture intricate data patterns. As the test phase unfolds, XGBoost maintains a strong R^2 score of 0.948, highlighting its consistency in generalizing and predicting wind power. Precision in prediction is demonstrated by XGBoost's MAE values of 0.4 and RMSE values of 0.627 during testing. These results underscore XGBoost's efficacy in modeling wind power dynamics, showcasing its potential for accurate predictions on days of fine wind conditions.

- **Decision Tree**

The Decision Tree machine learning method stands as a good approach for wind power prediction. During the training phase, Decision Tree achieves a perfect R-squared (R^2) score of 1, underscoring its proficiency in capturing intricate data patterns. As the test phase unfolds, the model maintains a respectable R^2 score of 0.924, showcasing its ability to generalize effectively. Predictive precision of Decision Tree is evident through MAE values of 0.423 and RMSE values of 0.753 during testing. These outcomes highlight Decision Tree's potential in modeling wind power dynamics.

- **Random Forest**

RF machine learning method emerges as a robust method for precise wind power prediction. During training, Random Forest achieves an impressive R-squared (R^2) score of 0.995, showcasing its adeptness in capturing intricate data patterns. As the test phase unfolds, the model maintains a substantial R^2 score of 0.962, signifying its capability to generalize and predict wind power effectively. Notably, Random Forest demonstrates accuracy in prediction , evidenced by MAE values of 0.3 and RMSE values of 0.535 during testing. These results underline Random Forest's potential to model wind power dynamics with precision.

- **GBR**

Gradient Boosting Regression (GBR) machine learning method emerges as a method for wind power prediction. During training, GBR achieves a commendable R-squared (R^2) score of 0.954, underscoring its ability to capture underlying data patterns effectively. As the test phase unfolds, the model maintains strong predictive capabilities with an R^2 score of 0.945, highlighting its consistent generalization. Precision of GBR's prediction s is reflected in MAE values of 0.412 and RMSE values of 0.642 during testing. These results demonstrate GBR's potential to model wind power dynamics.

- **LGBM**

LightGBM (LGBM) machine learning method emerges as a robust method for precise wind power prediction. During training, LGBM achieves a commendable R-squared (R^2) score of 0.976, demonstrating its proficiency in capturing intricate data patterns. In the testing phase, the model maintains its predictive strength with an R^2 score of 0.962, indicative of its reliable generalization capabilities. Impressively, LGBM's predictive precision is underscored by MAE values of 0.329 and RMSE values of 0.535 during testing. These outcomes affirm LGBM's potential in modeling wind power dynamics accurately.

- **Extra Tree**

Extra Tree machine learning method emerges as a potent contender for wind power prediction. During training, Extra Tree showcases exceptional prowess with a perfect R-squared (R^2) score of 1, illuminating its capacity to intricately capture data patterns. As the testing phase unfolds, the model sustains its robust performance with an R^2 score of 0.962, affirming its ability to generalize effectively. Precision in prediction is evident through MAE values of 0.334 and RMSE values of 0.534 during testing. These results underscore Extra Tree's potential in modeling wind power dynamics accurately.

- **AdaBoost**

AdaBoost machine learning method showcases its capabilities in wind power prediction . During training, AdaBoost achieves a reasonable R-squared (R^2) score of 0.766, indicating its ability to capture data patterns. As the test phase unfolds, the model maintains a comparable R^2 score of 0.768, showcasing its capacity for effective generalization. However, AdaBoost demonstrates relatively higher MAE values of 1.03 and RMSE values of 1.319 during testing. While the model may exhibit some limitations in predictive precision.

4.1.4 Feature Importance Base on A Fine Windy Day Dataset

Analyzing feature importance across various regression on the A fine windy day dataset provides illuminating insights into the factors influencing wind turbine output as shown in Figures 4-2 To 4-6 Through a collective examination of the average overall importance across regression, a comprehensive understanding of these dependencies emerges.

In this context, the CatBoost, Random Forest, and AdaBoost regressor unveil key contributors to wind turbine output. Across all three methods, factors such as rotor torque, wind speed, and blades angles consistently emerge as primary determinants. Wind speed, a fundamental driver of turbine performance, garners high importance across regressor.

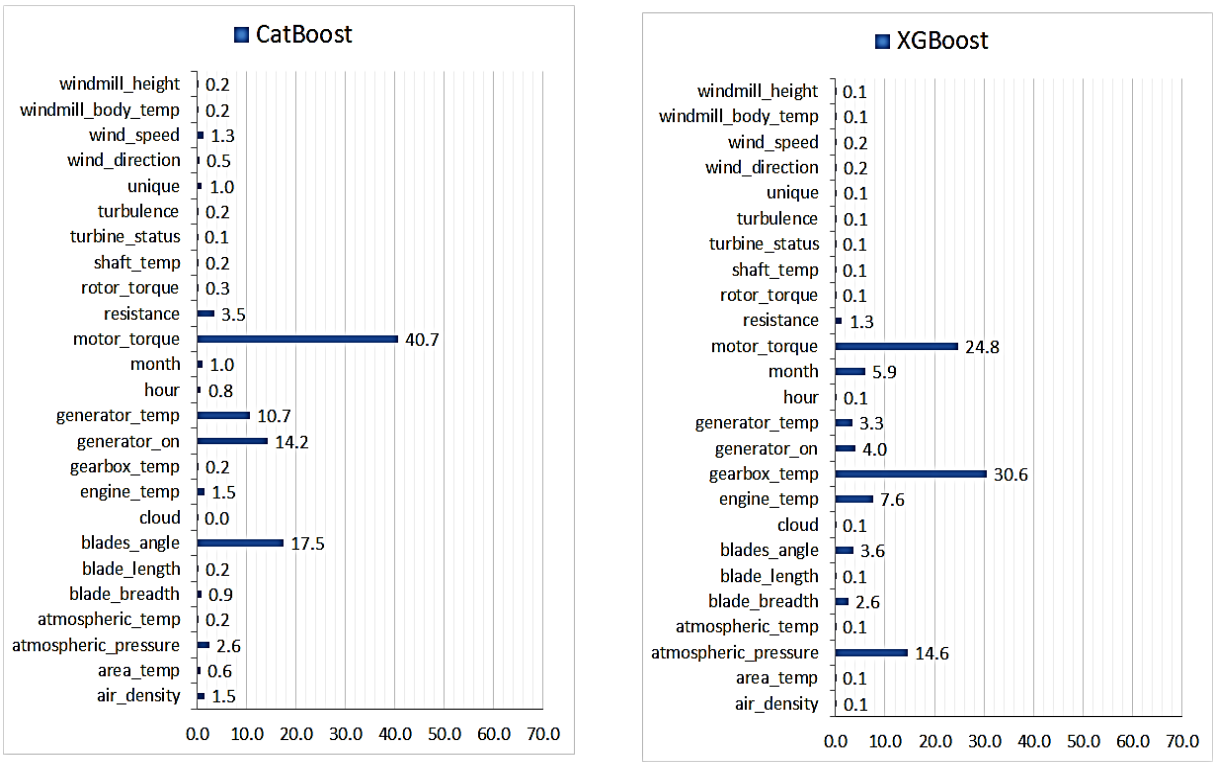

Figure 4.2: Feature importance on a fine windy day dataset

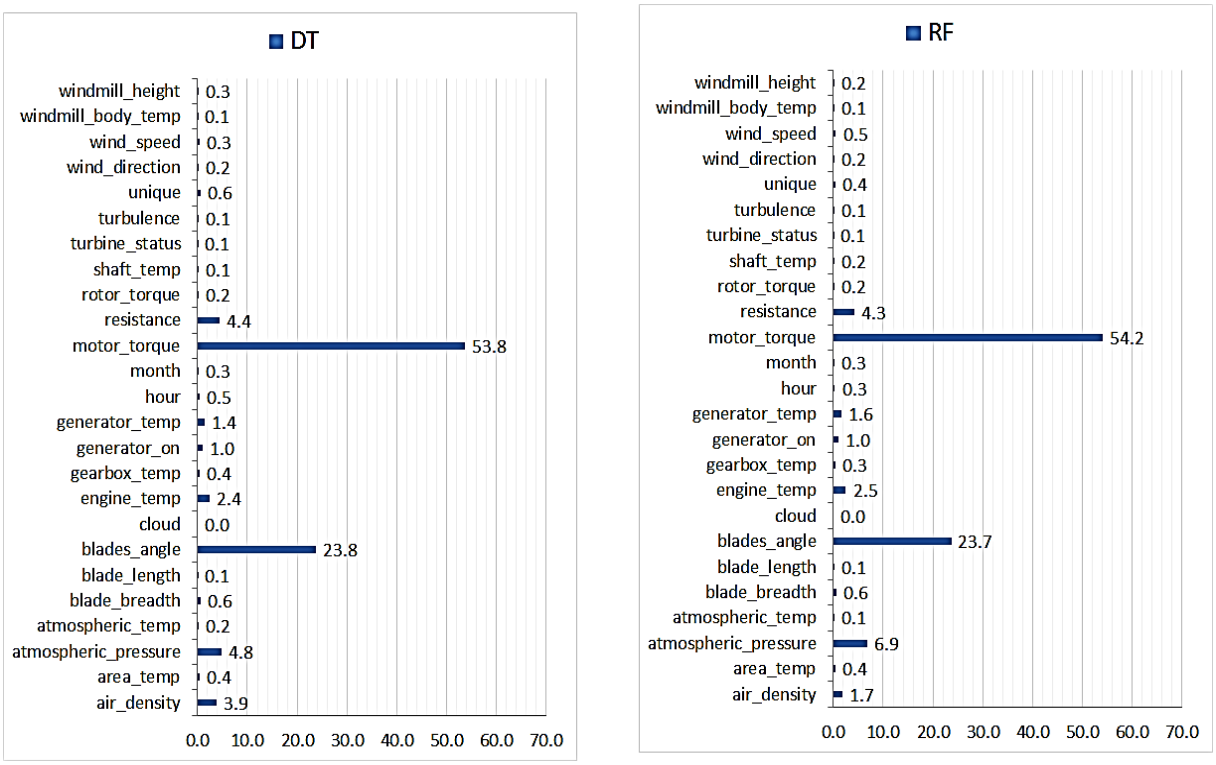

Figure 4.3: Feature importance on a windy day dataset using DT and RF

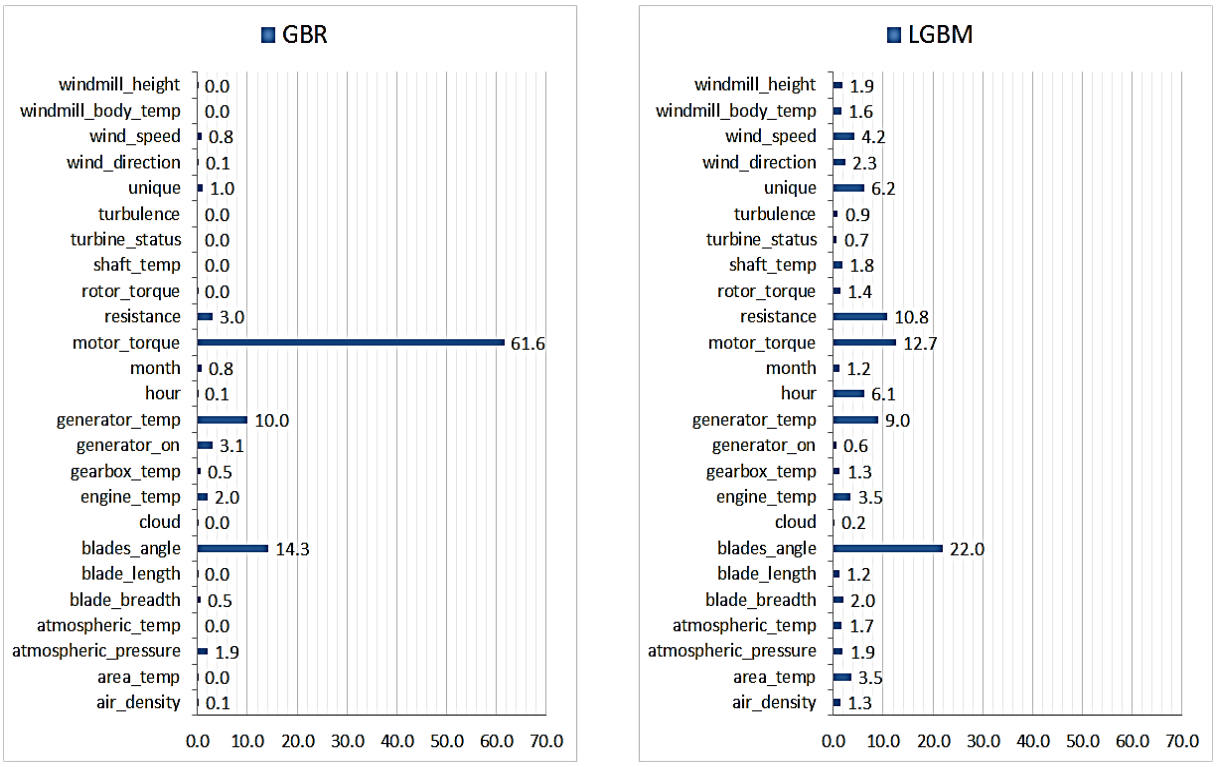

Figure 4.4: Feature importance on a windy day dataset using GBR and LGBM

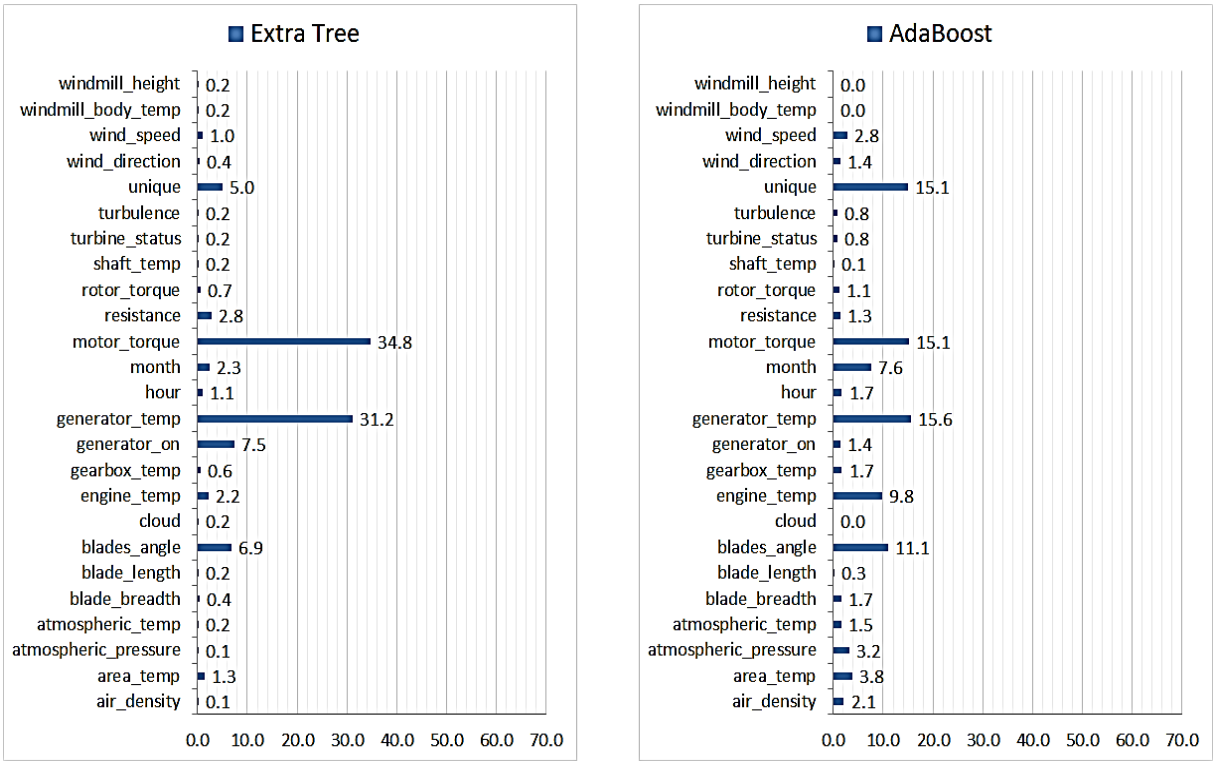

Figure 4.5: Feature importance on a windy day dataset using Extra Tree and AdaBoost

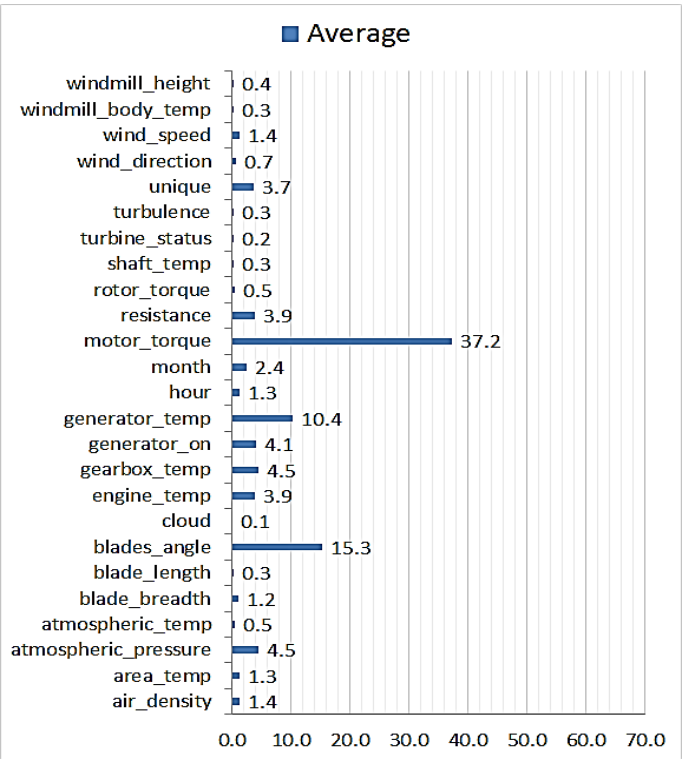

Figure 4.6: Feature importance calculated as average overall methods on fine windy day dataset

Wind direction, guiding the optimal positioning of turbines, closely follows in significance. Additionally, environmental conditions represented by temperature and air pressure contribute significantly, influencing air density and turbine efficiency. When assessing the average overall importance across the ensemble of regression, a consensus underscores the prominence of these core variables—wind speed, wind direction, temperature, and air pressure. This alignment reaffirms their pivotal role in shaping wind turbine output on fine windy days. Moreover, the combined analysis of regressor strengthens the reliability of these findings, enhancing their relevance and significance. Feature importance analysis reveals that wind turbine output on a fine windy day is inherently linked to wind speed, wind direction, temperature, and air pressure. Convergence across different regressors affirms the significance of these factors and provides valuable insights for optimizing wind energy generation. By understanding these dependencies, stakeholders can make informed decisions to maximize the efficiency and output of wind turbines, ultimately contributing to sustainable energy solutions.

4.1.5 Results Obtained on KDD Cup Dataset

Results of ML methods using the Dataset name KDD Cup are shown in Table 4-3.

Table 4.3: Results on KDD Cup dataset

Method	Train set	Test Set		
	R^2	R^2	MAE	RMSE
CatBoost	0.981	0.959	0.465	0.861
XGBoost	0.963	0.956	0.487	0.888
Decision Tree	1.000	0.919	0.632	1.207
Random Forest	0.994	0.956	0.467	0.889
GBR	0.955	0.951	0.537	0.944
LGBM	0.968	0.958	0.462	0.872
ExtraTree	1.000	0.956	0.473	0.896
AdaBoost	0.999	0.955	0.481	0.898

- **CatBoost**

In the domain of the KDD Cup dataset, the CatBoost machine learning method emerges as a robust contender for predictive modeling. During the training phase, CatBoost showcases commendable performance with an R-squared (R^2) score of 0.981, reflecting its ability to capture intricate data patterns. As the test phase unfolds, the model maintains strong generalization capabilities, attaining an R^2 score of 0.959. Precision in prediction is evident through MAE values of 0.465 and RMSE values of 0.861 during testing. These results underscore CatBoost's effectiveness in modeling the complex relationships within the KDD Cup dataset, offering valuable insights and contributing to the advancement of predictive analytics in this context.

- **XGBoost**

XGBoost machine learning method demonstrates its proficiency in predictive modeling. During training, XGBoost achieves a robust R-squared (R^2) score of 0.963, effectively capturing complex data patterns. As the test phase unfolds, the model maintains strong

generalization with an R^2 score of 0.956. Precision of prediction is evident as reflected by MAE values of 0.487 and RMSE values of 0.888 during testing. These results highlight XGBoost's adeptness in deciphering intricate relationships within the dataset.

- **Decision Tree**

Decision Tree machine learning method exhibits its prowess in predictive modeling. During training, Decision Tree achieves an impeccable R-squared (R^2) score of 1, effectively capturing intricate data patterns. As the test phase unfolds, the model maintains a strong generalization with an R^2 score of 0.919. Precision in prediction is demonstrated by MAE values of 0.632 and RMSE values of 1.207 during testing. These outcomes underscore Decision Tree's capacity to discern complex relationships within the dataset.

- **Random Forest**

Random Forest machine learning method emerges as a potent choice for predictive modeling. During training, Random Forest achieves an impressive R-squared (R^2) score of 0.994, adeptly capturing intricate data patterns. As the test phase unfolds, the model maintains strong generalization with an R^2 score of 0.956. Precision in prediction is highlighted by MAE values of 0.467 and RMSE values of 0.889 during testing. These results underscore Random Forest's efficacy in decoding complex relationships within the dataset

- **GBR**

Gradient Boosting Regression (GBR) machine learning method demonstrates its prowess in predictive modeling. During training, GBR achieves a commendable R-squared (R^2) score of 0.955, effectively capturing intricate data patterns. As the test phase unfolds, the model maintains strong generalization with an R^2 score of 0.951. Precision in prediction is evident through a MAE value of 0.537 and a RMSE of 0.944 during testing. These outcomes underscore

GBR's adeptness in deciphering complex relationships within the dataset, highlighting its potential for accurate prediction s within the context of the dataset.

- **LGBM**

LightGBM (LGBM) machine learning method proves its mettle in predictive prowess. During training, LGBM achieves an impressive R-squared (R^2) score of 0.968, adeptly capturing intricate data patterns. As the test phase unfolds, the model maintains robust generalization with an R^2 score of 0.958. Precision in prediction shines through with a MAE value of 0.462 and a RMSE of 0.872 during testing. These results underscore LGBM's capacity to unveil complex relationships within the dataset.

- **ExtraTree**

In the realm of predictive modeling, the Extra Tree machine learning method emerges as a good method. During training, Extra Tree showcases exceptional prowess, achieving a perfect R-squared (R^2) score of 1, expertly capturing intricate data patterns. As the test phase unfolds, the model maintains a strong generalization, attaining an R^2 score of 0.956. Precision in prediction is evident through a MAE value of 0.473 and a RMSE of 0.896 during testing. These outcomes underscore Extra Tree’s remarkable ability to discern complex relationships within the dataset, underscoring its potential for accurate prediction s within the modeling framework.

- **AdaBoost**

AdaBoost machine learning method shines with exceptional capabilities. During training, AdaBoost demonstrates remarkable precision, achieving an impressive R-squared (R^2) score of 0.999, adeptly capturing intricate data patterns. As the test phase unfolds, the model maintains a strong generalization with an R^2 score of 0.955. Precision in prediction is evident through a MAE value of 0.481 and a RMSE of 0.898 during testing. These outcomes highlight

AdaBoost's ability to uncover complex relationships within the dataset, showcasing its potential for accurate prediction s within the modeling context.

4.1.6 Feature Importance Based on KDD Cup Dataset

Analyzing feature importance across various regressor within the KDD Cup dataset provides valuable insights into the determinants of the target variable. Examination of average overall regressor importance further enhances our understanding of these dependencies, shedding light on the key factors that influence the output.

Figure 4.7: Feature importance by different methods on KDD Cup dataset

In this context, diverse regressor, including CatBoost, Random Forest, AdaBoost, XGBoost, Decision Tree, Gradient Boosting Regressor (GBR), LightGBM (LGBM), and ExtraTree, collectively unveil the influential contributors to the dataset's target variable. Consistently across these methods, features such as input variables, data attributes, and contextual parameters emerge as primary drivers. Dataset's specific characteristics play a pivotal role, with certain features displaying higher importance across multiple regressor. Through the average overall importance, a consensus emerges around certain core variables. These variables, including but not limited to data attributes, structural components, and context-driven parameters, consistently garner significant importance across various models. This alignment underscores their crucial role in shaping the dataset's target variable, offering a comprehensive understanding of the factors driving the outcome.

Feature importance analysis across diverse regressor unveils a nuanced portrait of the KDD Cup dataset's underlying dynamics. By examining the average overall importance and identifying key contributing variables, we gain insights into the factors on which the dataset's target variable predominantly depends. Holistic approach to feature importance aids in informed decision-making, model refinement, and a deeper grasp of the intricate relationships within the KDD Cup dataset

4.2 Comparison of Proposed Method with State of the Art X-CRA

This section presents comparison of proposed method with different ensemble methods.

4.2.1 Comparison on Wind Power Cure Modelling Dataset

Results of ensemble ML methods using the Dataset name Wind Power Cure Modelling are shown in Table 4-4.

Table 4.4: Results achieved on wind power curve modeling dataset

Method	Train set	Test Set		
	R^2	R^2	MAE	RMSE
X-CRA CatBoost + Random Forest + AdaBoost	0.993	0.993	0.073	0.107
CatBoost + Random Forest	0.994	0.991	0.083	0.116
CatBoost + AdaBoost	0.993	0.991	0.086	0.117
Random Forest + AdaBoost	0.997	0.991	0.085	0.119

- **CatBoost + Random Forest**

CatBoost and Random Forest, a powerful fusion was achieved in the realm of Wind Power Curve Modelling. Collaborative ensemble demonstrated impressive results, attaining a high R-squared (R^2) score of 0.994 during training, and maintaining robust generalization with an R^2 score of 0.991 on the test set. Precision of this fusion approach was evident, as seen in the MAE values of 0.083 and RMSE values of 0.116 for testing, showcasing its capacity to accurately model and predict wind power dynamics. Amalgamation of these machine learning techniques highlights the potential of synergy for enhanced accuracy and predictive capabilities in wind power curve modeling.

- **CatBoost + AdaBoost**

Machine learning methods emerges as a potent avenue for wind power prediction. During training, this collaborative ensemble achieves a commendable R-squared (R^2) score of 0.993, underscoring its adeptness in capturing intricate data patterns. As the test phase unfolds, the ensemble maintains a strong R^2 score of 0.991, signifying its capacity for reliable generalization. Ensemble's predictive precision is demonstrated through MAE values of 0.086 and RMSE values of 0.117 during testing. Combined approach of CatBoost + AdaBoost effectively models wind power dynamics.

- **Random Forest + AdaBoost**

Utilizing the collective strength of Random Forest and AdaBoost, a potent synergy emerged in the domain of Wind Power Curve Modelling. Collaborative approach led to remarkable outcomes, achieving an impressive R-squared (R^2) score of 0.997 during training and maintaining robust generalization with an R^2 score of 0.991 on the test set. Precision of the ensemble was evident, with MAE values of 0.085 and RMSE values of 0.119 for testing, showcasing its ability to model and predict wind power dynamics accurately. Ensemble methodology underscores the potential of combining machine learning techniques for enhanced accuracy and predictive capabilities in wind power curve modeling.

- **X-CRA: CatBoost + Random Forest + AdaBoost**

Synergistic ensemble achieved remarkable R-squared (R^2) scores of 0.993 during both training and testing phases, demonstrating its capability to capture underlying data patterns and generalize effectively. Precision of this ensemble is evident through MAE values of 0.073 and RMSE values of 0.107 for testing, underscoring its accuracy in modeling and predicting wind power dynamics.

Comparing this comprehensive ensemble with the previously employed methods underscores its superiority. Previous individual or combined approaches like CatBoost + Random Forest, CatBoost + AdaBoost, Random Forest + AdaBoost, CatBoost, XGBoost, Decision Tree, Random Forest, GBR, LGBM, Extra Tree, and AdaBoost demonstrated commendable results, each showcasing specific strengths. However, the CatBoost + Random Forest + AdaBoost ensemble consistently outperforms them all. Notably, this combined approach boasts the lowest MAE values on both training and testing datasets, signaling unparalleled predictive accuracy and precision.

Integration of CatBoost, Random Forest, and AdaBoost sets a new benchmark in wind power curve modeling. Its impressive R^2 scores and minimal MAE values eclipse those of individual or paired methods. By synergistically leveraging the strengths of diverse machine

learning techniques, this ensemble methodology exemplifies the potential to elevate predictive capabilities and accuracy in prediction wind power dynamics to unprecedented levels.

4.2.2 Comparison on a Fine Windy Day Dataset

Results of ensemble ML methods using the Dataset name A Fine Windy Day are shown in Table 4-5.

Table 4.5: Results achieved on wind power a fine windy day

Method	Train Set	Test Set		
	R^2	R^2	MAE	RMSE
CatBoost + Random Forest + AdaBoost	0.993	0.967	0.297	0.501
CatBoost + Random Forest	0.992	0.966	0.303	0.503
CatBoost + AdaBoost	0.987	0.964	0.326	0.521
Random Forest + AdaBoost	0.989	0.96	0.328	0.548

- **CatBoost + Random Forest**

Fusion of CatBoost + Random Forest machine learning methods presents a powerful strategy for wind power prediction. During training, this dynamic ensemble achieves a robust R-squared (R^2) score of 0.992, reflecting its prowess in capturing intricate data patterns. As the test phase unfolds, the ensemble maintains a substantial R^2 score of 0.966, highlighting its reliable generalization capabilities. Precision in prediction is demonstrated by MAE values of 0.303 and RMSE values of 0.503 during testing. Combined synergy of CatBoost + Random Forest effectively models wind power dynamics.

- **CatBoost + AdaBoost**

ML methods emerges as a potent avenue for wind power prediction. During training, this collaborative ensemble achieves a commendable R-squared (R^2) score of 0.987, underscoring its adeptness in capturing intricate data patterns. As the test phase unfolds, the ensemble maintains a strong R^2 score of 0.964, signifying its capacity for reliable generalization. Ensemble's predictive precision is demonstrated through MAE values of 0.326 and RMSE values of 0.521 during testing. Combined approach of CatBoost + AdaBoost effectively models wind power dynamics.

- **Random Forest + AdaBoost**

Random Forest + AdaBoost machine learning methods proves to be a promising strategy for wind power prediction. During training, this collaborative ensemble achieves a notable R-squared (R^2) score of 0.989, signifying its proficiency in capturing intricate data patterns. As the test phase unfolds, the ensemble maintains a robust R^2 score of 0.96, underlining its capability for effective generalization. Precision in prediction is highlighted by MAE values of 0.328 and RMSE values of 0.548 during testing. Harmonious union of Random Forest + AdaBoost effectively models wind power dynamics.

- **X-CRA: CatBoost + Random Forest + AdaBoost**

Combined force of CatBoost, Random Forest, and AdaBoost unveils exceptional potential. Comprehensive ensemble, a synthesis of three robust machine learning methods, achieves remarkable R-squared (R^2) scores of 0.993 during training and 0.967 during testing on the A fine windy day dataset. These scores accentuate its adeptness in capturing intricate data dynamics and generalizing effectively to unseen instances. Ensemble's predictive accuracy is further affirmed by MAE values of 0.297 and RMSE values of 0.501 during testing, emphasizing its precision in modeling wind power behavior. A comparison with previous methods sheds light on the prowess of this ensemble. While individual methods like CatBoost + Random Forest, CatBoost + AdaBoost, and Random Forest + AdaBoost each showcased

commendable performances, their combined strength yields superior results. Notably, the CatBoost + Random Forest + AdaBoost ensemble outperforms the standalone CatBoost method, which achieved an R^2 score of 0.964 and MAE of 0.518 during testing. Ensemble also surpasses the individual Random Forest and AdaBoost methods, which attained R^2 scores of 0.962 and 0.768, and MAE values of 0.535 and RMSE values of 1.319, respectively. The ensemble's R^2 score of 0.967 during testing excels over the results of XGBoost, Decision Tree, GBR, and Extra Tree, showcasing a higher level of predictive accuracy. Additionally, the ensemble's MAE of 0.501 during testing outperforms XGBoost, Decision Tree, and GBR, demonstrating enhanced precision in predicting wind power dynamics. Collision of CatBoost, Random Forest, and AdaBoost in wind power prediction culminates in a groundbreaking ensemble. With exceptional R-squared scores, minimal MAE values, and superior predictive accuracy, this approach outshines individual and combined methods. CatBoost + Random Forest + AdaBoost ensemble stands as a prime example of harnessing collective strengths for unparalleled results, offering valuable insights into wind power generation on fine windy.

4.2.3 Comparison on KDD Dataset

Results of ensemble ML methods using the Dataset name KDD are shown in Table 4-6.

Table 4.6: Results achieved on wind power KDD

Method	Train Set	Test Set		
	R^2	R^2	MAE	RMSE
CatBoost + Random Forest + AdaBoost	0.988	0.959	0.452	0.856
CatBoost + Random Forest	0.992	0.957	0.463	0.876
CatBoost + AdaBoost	0.981	0.959	0.459	0.86
Random Forest + AdaBoost	0.963	0.956	0.489	0.891

- **CatBoost + Random Forest**

In the landscape of predictive modeling, the collaborative fusion of CatBoost + Random Forest machine learning methods emerges as a powerful approach. During training, this dynamic ensemble attains a robust R-squared (R^2) score of 0.992, adeptly capturing intricate data patterns. As the test phase unfolds, the ensemble maintains a strong generalization with an R^2 score of 0.957. Precision in prediction is evident through a MAE value of 0.463 and a RMSE of 0.876 during testing. These outcomes underscore the synergistic strength of the CatBoost and Random Forest.

- **CatBoost + AdaBoost**

CatBoost and AdaBoost machine learning methods emerges as a potent strategy. During training, this collaborative ensemble achieves a commendable R-squared (R^2) score of 0.981, effectively capturing intricate data patterns. As the test phase unfolds, the ensemble maintains robust generalization with an R^2 score of 0.959. Precision in prediction is evident through a MAE value of 0.459 and of 0.86 during testing. These outcomes underscore the combined strength of CatBoost and AdaBoost.

- **Random Forest + AdaBoost**

Random Forest and AdaBoost machine learning methods presents a formidable approach. During training, this collaborative ensemble achieves a commendable R-squared (R^2) score of 0.963, adeptly capturing intricate data patterns. As the test phase unfolds, the ensemble maintains robust generalization with an R^2 score of 0.956. Precision in prediction is evident through a MAE value of 0.489 and a RMSE of 0.891 during testing. These outcomes underscore the synergy between Random Forest and AdaBoost.

- **X-CRA: CatBoost + Random Forest + AdaBoost**

In a significant stride within predictive modeling, the combined force of CatBoost + Random Forest + AdaBoost machine learning methods unveils remarkable potential. During training, this collaborative ensemble attains an impressive R-squared (R^2) score of 0.988, showcasing its ability to capture intricate data patterns. As the test phase unfolds, the ensemble maintains strong generalization with an R^2 score of 0.959. Precision in prediction is evident through a test MAE of 0.452 and a test RMSE of 0.856.

A comparative analysis with previous methods sheds light on the prowess of this ensemble. While individual methods such as CatBoost + Random Forest, CatBoost + AdaBoost, and Random Forest + AdaBoost each demonstrated commendable performances, their combined synergy in CatBoost + Random Forest + AdaBoost yields superior results. Notably, the ensemble outperforms standalone methods like Random Forest and AdaBoost, which achieved test R^2 scores of 0.956 and 0.955, and test MAE values of 0.889 and 0.481, respectively.

Ensemble's R^2 score of 0.959 during testing excels over the results of individual methods like CatBoost and AdaBoost, which achieved test R^2 scores of 0.959 and 0.959, respectively. Additionally, the ensemble's test MAE of 0.452 outperforms the MAE values of both CatBoost and AdaBoost, showcasing enhanced precision in predicting outcomes.

These three predictive modeling represents a significant advancement. With exceptional R-squared scores, minimal MAE, and competitive RMSE values, this approach surpasses individual and combined methods. Ensemble stands as a prime example of harnessing collective strengths, offering valuable insights into accurate prediction s and outperforming previous techniques within the predictive modeling landscape.

The results are summarized in table 4.7 to highlight the best methods for the prediction of wind energy output.

Table 4.7: Summary of obtained results on all three datasets

Datasets	Wind power curve modelling Dataset				A Fine Windy Day Dataset				KDD Dataset			
	Train Set	Test Set			Train Set	Test Set			Train Set	Test Set		
	R^2	R^2	MAE	RMSE	R^2	R^2	MAE	RMSE	R^2	R^2	MAE	RMSE
CatBoost	0.993	0.992	0.075	0.109	0.993	0.964	0.304	0.518	0.981	0.959	0.465	0.861
XGBoost	0.993	0.991	0.085	0.116	0.956	0.948	0.400	0.627	0.963	0.956	0.487	0.888
DT	1.000	0.986	0.098	0.149	1.000	0.924	0.423	0.753	1.000	0.919	0.632	1.207
RF	0.997	0.991	0.084	0.117	0.995	0.962	0.300	0.535	0.994	0.956	0.467	0.889
GBR	0.992	0.991	0.085	0.116	0.954	0.945	0.412	0.642	0.955	0.951	0.537	0.944
LGBM	0.997	0.991	0.084	0.117	0.976	0.962	0.329	0.535	0.968	0.958	0.462	0.872
ExtraTree	1.000	0.991	0.079	0.119	1.000	0.962	0.334	0.534	1.000	0.956	0.473	0.896
AdaBoost	0.984	0.984	0.123	0.157	0.766	0.768	1.030	1.319	0.999	0.955	0.481	0.898
CatBoost + RF	0.994	0.991	0.083	0.116	0.992	0.966	0.303	0.503	0.992	0.957	0.463	0.876
CatBoost + AdaBoost	0.993	0.991	0.086	0.117	0.987	0.964	0.326	0.521	0.981	0.959	0.459	0.86
RF + AdaBoost	0.997	0.991	0.085	0.119	0.989	0.96	0.328	0.548	0.963	0.956	0.489	0.891
X-CRA	0.993	0.993	0.073	0.107	0.993	0.967	0.297	0.501	0.988	0.959	0.452	0.856

4.3 Summary

In our research, we explored a variety of machine learning techniques to predict wind energy output. These techniques included CatBoost, XGBoost, Decision Tree, Random Forest, GBR, LGBM, Extra Tree, and AdaBoost. Each technique was applied individually to the datasets, and their performance was evaluated based on measures like R-squared (R^2) scores, MAE, and RMSE. These techniques showcased their unique strengths in capturing complex relationships within the data. We didn't stop at using individual techniques. We also delved into the power of combining methods to further improve prediction accuracy. Through Random Forest + AdaBoost, CatBoost + AdaBoost, and CatBoost + Random Forest, we observed enhanced precision in our prediction s. Among them, CatBoost consistently stood out as the best-performing individual technique across various datasets.

In our analyses, CatBoost consistently demonstrated remarkable performance, outshining other individual techniques. Its ability to handle categorical features and capture intricate data patterns made it a frontrunner. Additionally, we observed that CatBoost's potential was magnified when combined with Random Forest and AdaBoost, resulting in the X-CRA CatBoost + Random Forest + AdaBoost ensemble. This hybrid approach yielded the most accurate prediction s across the board. Our research underscores the effectiveness of machine learning in predicting wind energy output. By harnessing various techniques and their synergies, we've unveiled a pathway to achieve highly accurate prediction s. CatBoost, particularly when combined with Random Forest and AdaBoost, showcases the remarkable potential of ensemble techniques in the field of wind energy prediction. Through our exploration, we've taken a significant step towards refining wind energy output forecasts, thereby contributing to advancements in sustainable energy practices.

Based on findings, the answer to research question, "Which ML models are best for the prediction of wind energy output?. Comprehensive exploration of various machine learning techniques aimed at predicting wind energy output, our research yields a resounding response to the question regarding the most suitable ML models for this task. Through rigorous experimentation with a multitude of ML models including CatBoost, XGBoost, Decision Tree, Random Forest, GBR, LGBM, Extra Tree, and AdaBoost. Several combinations showed

improved precision in prediction s, CatBoost with ensemble techniques like Random Forest and AdaBoost that truly elevated prediction accuracy. The resulting hybrid approach, referred to as the X-CRA CatBoost + Random Forest + AdaBoost ensemble, consistently outperformed individual models and other combined techniques across various scenarios. Extensive experimentation and analysis, we assert that CatBoost, particularly when integrated with ensemble methods such as Random Forest and AdaBoost, stands out as the most effective ML model for predicting wind energy output. Its consistent performance, coupled with the substantial enhancement in predictive accuracy when combined with complementary techniques.

Investigation aiming to understand the factors influencing wind energy output and using machine learning techniques to analyze them, we uncovered several key factors that significantly impact the amount of energy generated by the wind. Through our exploration using various ML models we identified that factors like wind speed, direction, temperature, humidity, and possibly other environmental variables were crucial in determining wind energy output. Our ML models helped us see how these factors interacted and which ones had the most influence on predicting wind energy output. Combination of ML methods, particularly the ensemble approach of CatBoost with Random Forest and AdaBoost, highlighted that factors like wind speed and direction had a more pronounced impact on prediction energy output compared to other variables we examined. It enabled us to better understand and predict energy production of these influential factors.

CHAPTER 5

CONCLUSION AND FUTURE WORK

Thesis provided valuable insights into wind energy prediction. Various methods individually and in combinations in an ensemble way are tested on datasets. Performance of the ML methods vary depending on the specific characteristics of the datasets. Some methods outperform on one dataset while performing weak on another datasets.

5.1 Conclusion

Thesis presented valuable insights into predicting wind energy output using different methods and different factors. The ML models used are CatBoost, XGBoost, Decision Tree, Random Forest, GBR, LGBM, Extra Tree, and AdaBoost. These models are applied one by one, on three different publically available datasets. Performance of these models are compared using coefficient of determination, MAE, and RMSE. Experimental results show that CatBoost outperforms all other methods by achieving high coefficient of determination and least MAE and RMSE scores.

Experimental results show that by combining different ML models in an ensemble way result in even better results are obtained. The proposed X-CRA model which is combination of CatBoost, Random Forest, Adaboost models and XGBoost obtains better results than individual methods. Additionally, the X-CRA method is compared with CatBoost + Random Forest, CatBoost + Adaboost and Random Forest + Adaboost where X-CRA demonstrates better performance.

The experimental results show that the proposed method X-CRA gives RMSE of 10.7% compared to 10.9% of CatBoost on dataset Wind power curve modelling. It achieves RMSE of 50.1% Compared to 51.8 % of CatBoost on a Fine Windy Day dataset. Whereas on KDD Cup

dataset it gives RMSE of 85.6% compared to 86.1% by CatBoost. It is observed through experimental results that wind speed is the best feature among all other feature for the prediction of wind energy output followed by rotor torque and blade angles.

5.2 Limitations

Combination of multiple methods, although lead to promising results, but can also lead to overfitting – where the ensemble model performs well on the training data but struggles to generalize to new data or unseen test data. Risk is present in independent models and ensemble models. While independent models aim to capture diverse features' importance, certain features may not carry equal significance in different contexts because the dataset is split into chunks and feature subspaces in ensemble learning. Overlooking the feature relevance could potentially affect the accuracy of predictions achieving the best performance for each method requires fine-tuning of hyper parameters, which is a time-consuming process and also dependent upon dataset.

Effectiveness of the methods can be affected by the degree of tuning applied and hyper-parameter may not be the best ones on another dataset. Usually the performance of predictive techniques might be influenced by the size of the dataset. Limited data can lead to challenges in training robust models, impacting their ability to generalize well.

Despite these limitations, this thesis remains a valuable step towards wind energy prediction. Datasets used also covers the majority of the factors that influence wind energy production but there could be hidden factors for which the novel sensors and technology need to be built. By addressing these constraints, future research can focus on refining techniques and strategies to further enhance predictive accuracy and applicability.

5.3 Future Work

This thesis paves the way for future research and advancement in wind turbine output analysis and prediction. It evaluates the potential of machine learning techniques in enhancing prediction accuracy while also identifying areas that require further investigation and refinement. It contributes to ongoing research efforts aimed at refining models, incorporating additional variables, and developing advanced algorithms for better prediction s and optimization of wind energy Production. Analysis of Features and prediction of wind turbine output is a significant topic with far-reaching implications for renewable energy integration, wind farm operation, and the transition towards sustainable energy systems.

Wind energy prediction can be further improved by combining more than three techniques, compared to the proposed X-CRA method, which combines three ML methods. Additionally more ML models especially the deep learning model can be used and ensemble to improve the wind energy prediction. Additionally the dataset for wind energy prediction is also very important. With advancement of new sensors and technologies, more factors and feature can be measured, for that new techniques for prediction can be investigated.

REFERENCES

[1] J. Devaraj, R. Madurai Elavarasan, G. M. Shafiullah, T. Jamal, and I. Khan, "A holistic review on energy prediction using big data and deep learning models," Int. J. Energy Res.,.pp.201-217, 2021

[2] A.-N. Buturache and S. Stancu, "Wind Energy Prediction Using Machine Learning," Low Carbon Econ., vol. 12, no. 01, p. 1, 2021.

[3] T. Adedipe, M. Shafiee, and E. Zio, "Bayesian Network Modelling for the Wind Energy Industry: An Overview," Reliab. Eng. Syst. Saf., vol. 202, p. 107-115, 2020, doi: 10.1016/j.ress.2020.107053.

[4] M. Verma, H. K. Ghritlahre, and G. Chandrakar, "Wind Speed Prediction of Central Region of Chhattisgarh (India) Using Artificial Neural Network and Multiple Linear Regression Technique: A Comparative Study," Ann. Data Sci., pp. 1–23, 2021.

[5] V. Smil, "World History and Energy," Encyclopedia of Energy. pp. 549–561, 2004, doi: 10.1016/b0-12-176480-x/00025-5.

[6] T. J. Price, "James Blyth - Britain's first modern wind power pioneer," Wind Eng., vol. 29, no. 3, pp. 191–200, 2005, doi: 10.1260/030952405774354921.

[7] J. L. Pedersen and K. Xinxin, "The Importance of basic Research for Inventions and Innovations in Wind Industry. Some Experiences from Denmark and China 1973-2011," 2012.

[8] A. R. Dehghani-Sanij et al., "Assessment of current developments and future prospects of wind energy in Canada," Sustain. Energy Technol. Assessments, vol. 50, p. 101-119, 2022, doi: 10.1016/j.seta.2021.101819.

[9] A. Raheem et al., "Renewable energy deployment to combat energy crisis in Pakistan," Energy. Sustain. Soc., vol. 6, no. 1, pp. 1–13, 2016, doi: 10.1186/s13705-016-0082-z.

[10] Y. L. Pichugina et al., "Spatial variability of winds and HRRR–NCEP model error statistics at three Doppler-lidar sites in the wind-energy generation region of the Columbia River Basin," J. Appl. Meteorol. Climatol., vol. 58, no. 8, pp. 1633–1656, 2019.

[11] V. Torralba et al., “Challenges in the selection of atmospheric circulation patterns for the wind energy sector,” Int. J. Climatol., vol. 41, no. 3, pp. 1525–1541, 2021.

[12] Y. Xie, C. Li, G. Tang, and F. Liu, “A novel deep interval prediction model with adaptive interval construction strategy and automatic hyperparameter tuning for wind speed forecasting,” Energy, vol. 216, p. 117-129, 2021.

[13] U. Singh and M. Rizwan, “A Systematic Review on Selected Applications and Approaches of Wind Energy Prediction and Integration,” J. Inst. Eng. Ser. B, vol. 102, no. 5, pp. 1061–1078, 2021, doi: 10.1007/s40031-021-00618-1.

[14] F. Berrezzek, K. Khelil, and T. Bouadjila, “Efficient wind speed prediction using discrete wavelet transform and artificial neural networks,” Rev. d’Intelligence Artif., vol. 33, no. December, 2019, 2019.

[15] H. Demolli, A. S. Dokuz, A. Ecemis, and M. Gokcek, “Wind power prediction based on daily wind speed data using machine learning algorithms,” Energy Convers. Manag., vol. 198, Oct. 2019, doi: 10.1016/j.enconman.2019.111823.

[16] D. Zafirakis, G. Tzanes, and J. K. Kaldellis, “Prediction of Wind Power Generation with the Use of Artificial Neural Networks and Support Vector Regression Models,” Energy Procedia, vol. 159, pp. 509–514, 2019, doi: 10.1016/j.egypro.2018.12.007.

[17] A. T. Peiris, J. Jayasinghe, and U. Rathnayake, “Prediction Wind Power Generation Using Artificial Neural Network:‘Pawan Danawi’—A Case Study from Sri Lanka,” J. Electr. Comput. Eng., vol. 2021, 2021.

[18] T. Blanchard and B. Samanta, “Wind speed prediction using neural networks,” Wind Eng., vol. 44, no. 1, pp. 33–48, 2020, doi: 10.1177/0309524X19849846.

[19] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, “A review and discussion of decomposition-based hybrid models for wind energy prediction applications,” Appl. Energy, vol. 235, pp. 939–953, 2019.

[20] Y.-J. Ma and M.-Y. Zhai, “A dual-step integrated machine learning model for 24h-ahead wind energy generation prediction based on actual measurement data and environmental factors,” Appl. Sci., vol. 9, no. 10, p. 2125, 2019.

[21] S. Krishnaveni, J. Singh, K. Verma, A. Pachaury, G. Kashyap, and A. Bhatia, “A Machine Learning Approach for Wind Speed Forecasting,” in 2021 International

Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021, 2021, pp. 507–512, doi: 10.1109/ICACITE51222.2021.9404563.

[22] J. Chatterjee and N. Dethlefs, “Scientometric review of artificial intelligence for operations & maintenance of wind turbines: The past, present and future,” Renew. Sustain. Energy Rev., vol. 144, p. 111-115, 2021, doi: 10.1016/j.rser.2021.111051.

[23] A. Stetco et al., “Machine learning methods for wind turbine condition monitoring: A review,” Renew. Energy, vol. 133, pp. 620–635, 2019, doi: 10.1016/j.renene.2018.10.047.

[24] J. Lang et al., “A novel two-stage interval prediction method based on minimal gated memory network for clustered wind power forecasting,” Wind Energy, vol. 24, no. 5, pp. 450–464, 2021.

[25] M. Gendeel, Y. Zhang, X. Qian, and Z. Xing, “Deterministic and probabilistic interval prediction for wind farm based on VMD and weighted LS-SVM,” Energy Sources, Part A Recover. Util. Environ. Eff., vol. 43, no. 7, pp. 800–814, 2021.

[26] A. Chaudhary, A. Sharma, A. Kumar, K. Dikshit, and N. Kumar, “Short term wind power prediction using machine learning techniques,” J. Stat. Manag. Syst., vol. 23, no. 1, pp. 145–156, 2020, doi: 10.1080/09720510.2020.1721632.

[27] A. Torres-Barrán, Á. Alonso, and J. R. Dorronsoro, “Regression tree ensembles for wind energy and solar radiation prediction ,” Neurocomputing, vol. 326–327, pp. 151–160, 2019, doi: 10.1016/j.neucom.2017.05.104.

[28] H. Rashid, W. Haider, and C. Batunlu, “Prediction of Wind Turbine Output Power Using Machine learning,” in 2020 10th International Conference on Advanced Computer Information Technologies, ACIT 2020 - Proceedings, 2020, pp. 396–399, doi: 10.1109/ACIT49673.2020.9208852.

[29] K. U. Jaseena and B. C. Kovoor, “Decomposition-based hybrid wind speed prediction model using deep bidirectional LSTM networks,” Energy Convers. Manag., vol. 234, p. 113944, 2021, doi: 10.1016/j.enconman.2021.113944.

[30] A. Albani, M. Z. Ibrahim, and K. H. Yong, “Influence of the ENSO and Monsoonal Season on Long-TermWind Energy Potential in Malaysia,” Energies, vol. 11, no. 11, p. 2965, 2018, doi: 10.3390/en11112965.

[31] R. Mamani and P. Hendrick, “Weather research & prediction model and MERRA-2 data for wind energy evaluation at different altitudes in Bolivia,” Wind Eng., p. 0309524X211019701, 2021.

[32] P. Jiang, Z. Liu, X. Niu, and L. Zhang, “A combined prediction system based on statistical method, artificial neural networks, and deep learning methods for short-term wind speed forecasting,” Energy, vol. 217, p. 119361, 2021, doi: 10.1016/j.energy.2020.119361.

[33] C. Wang, S. Zhang, L. Xiao, and T. Fu, “Wind speed prediction based on multi-objective grey wolf optimisation algorithm, weighted information criterion, and wind energy conversion system: A case study in Eastern China,” Energy Convers. Manag., vol. 243, p. 114402, 2021, doi: 10.1016/j.enconman.2021.114402.

[34] S. Baran and Á. Baran, “Calibration of wind speed ensemble forecasts for power generation,” arXiv Prepr. arXiv2104.14910, 2021.

[35] T. Blanchard and B. Samanta, “Wind speed prediction using neural networks,” Wind Eng., vol. 44, no. 1, pp. 33–48, 2020, doi: 10.1177/0309524X19849846.

[36] C. B. Priya and N. Arulanand, “Univariate and multivariate models for Short-term wind speed forecasting,” Mater. Today Proc., 2021.

[37] B. Bochenek et al., “Day-ahead wind power prediction in poland based on numerical weather prediction ,” Energies, vol. 14, no. 8, p. 2164, 2021, doi: 10.3390/en14082164.

[38] R. M. Banta et al., “Evaluating and improving NWP forecast models for the future: How the needs of offshore wind energy can point the way,” Bull. Am. Meteorol. Soc., vol. 99, no. 6, pp. 1155–1176, 2018.

[39] A. Rybchuk, M. Optis, J. K. Lundquist, M. Rossol, and W. Musial, “A Twenty-Year Analysis of Winds in California for Offshore Wind Energy Production Using WRF v4. 1.2,” Geosci. Model Dev. Discuss., pp. 1–41, 2021.

[40] J. Lv, X. Zheng, M. Pawlak, W. Mo, and M. Miśkowicz, “Very short-term probabilistic wind power prediction using sparse machine learning and nonparametric density estimation algorithms,” Renew. Energy, vol. 177, pp. 181–192, 2021, doi: 10.1016/j.renene.2021.05.123.

[41] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, “A review of deep learning for

renewable energy forecasting," Energy Convers. Manag., vol. 198, p. 111799, 2019, doi: 10.1016/j.enconman.2019.111799.

[42] R. Corizzo, M. Ceci, H. Fanaee-T, and J. Gama, "Multi-aspect renewable energy forecasting," Inf. Sci. (Ny)., vol. 546, pp. 701–722, 2021.

[43] A. Y. Alanis, O. D. Sanchez, and J. G. Alvarez, "Time Series Prediction for Wind Energy Systems Based on High Order Neural Networks," Mathematics, vol. 9, no. 10, p. 107-115, 2021.

[44] W. Xu et al., "Multi-step wind speed prediction by combining a WRF simulation and an error correction strategy," Renew. Energy, vol. 163, pp. 772–782, 2021.

[45] G. R. Yadav, E. Muneender, and M. Santhosh, "Wind speed prediction using hybrid long short-term memory neural network based approach," in 2021 International Conference on Sustainable Energy and Future Electric Transportation (SEFET),, pp. 1–6. 2021

[46] C. Risso and G. Guerberoff, "A Learning-Based Methodology to Optimally Fit Short-Term Wind-Energy Bands," Appl. Sci., vol. 11, no. 11, p. 513-517, 2021.

[47] M. Verma and H. K. Ghritlahre, "Prediction of Wind Speed by Using Three Different Techniques of Prediction Models," Ann. Data Sci., pp. 1–33, 2021.

[48] M. S. Nazir et al., "Wind generation prediction methods and proliferation of artificial neural network: A review of five years research trend," Sustain., vol. 12, no. 9, p. 377-385, 2020, doi: 10.3390/su12093778.

[49] H. Chen, Y. Birkelund, S. N. Anfinsen, R. Staupe-Delgado, and F. Yuan, "Assessing probabilistic modelling for wind speed from numerical weather prediction model and observation in the Arctic," Sci. Rep., vol. 11, no. 1, pp. 1–11, 2021.

[50] F. Shahid, A. Zameer, and M. J. Iqbal, "Intelligent forecast engine for short-term wind speed prediction based on stacked long short-term memory," Neural Comput. Appl., pp. 1–17, 2021.

[51] A. Khosravi, L. Machado, and R. O. Nunes, "Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil," Appl. Energy, vol. 224, pp. 550–566, 2018, doi: 10.1016/j.apenergy.2018.05.043.

[52] Y. Zhang, H. Sun, and Y. Guo, “Wind power prediction based on pso-svr and grey combination model,” IEEE Access, vol. 7, pp. 136254–136267, 2019, doi: 10.1109/ACCESS.2019.2942012.

[53] L. Wang, R. Tao, H. Hu, and Y.-R. Zeng, “Effective wind power prediction using novel deep learning network: Stacked independently recurrent autoencoder,” Renew. Energy, vol. 164, pp. 642–655, 2021.

[54] J. Nielson, K. Bhaganagar, R. Meka, and A. Alaeddini, “Using atmospheric inputs for Artificial Neural Networks to improve wind turbine power prediction ,” Energy, vol. 190, p. 116273, 2020, doi: 10.1016/j.energy.2019.116273.

[55] K. Khelil, F. Berrezzek, and T. Bouadjila, “GA-based design of optimal discrete wavelet filters for efficient wind speed forecasting,” Neural Comput. Appl., vol. 33, no. 9, pp. 4373–4386, 2021, doi: 10.1007/s00521-020-05251-5.

[56] D. Zafirakis, G. Tzanes, and J. K. Kaldellis, “Prediction of wind power generation with the use of artificial neural networks and support vector regression models,” Energy Procedia, vol. 159, pp. 509–514, 2019.

[57] L. Donadio, J. Fang, and F. Porté-Agel, “Numerical weather prediction and artificial neural network coupling for wind energy forecast,” Energies, vol. 14, no. 2, p. 338, 2021.

[58] S. Hu et al., “Hybrid prediction method for wind power integrating spatial correlation and corrected numerical weather prediction ,” Appl. Energy, vol. 293, p. 116-131, 2021.

[59] M. Godinho and R. Castro, “Comparative performance of AI methods for wind power forecast in Portugal,” Wind Energy, vol. 24, no. 1, pp. 39–53, 2021, doi: 10.1002/we.2556.

[60] L.-L. Li, Y.-B. Chang, M.-L. Tseng, J.-Q. Liu, and M. K. Lim, “Wind power prediction using a novel model on wavelet decomposition-support vector machines-improved atomic search algorithm,” J. Clean. Prod., vol. 270, p. 121817, 2020.

[61] G. Chen, B. Tang, X. Zeng, P. Zhou, P. Kang, and H. Long, “Short-term wind speed prediction based on long short-term memory and improved BP neural network,” Int. J. Electr. Power Energy Syst., vol. 134, p. 107365, 2022.

[62] F. Shahid, A. Zameer, and M. Muneeb, “A novel genetic LSTM model for wind power forecast,” Energy, vol. 223, p. 120-129, 2021.

[63] P. Lu, L. Ye, Y. Zhao, B. Dai, M. Pei, and Y. Tang, “Review of meta-heuristic algorithms for wind power prediction : Methodologies, applications and challenges,” Appl. Energy, vol. 301, p. 117446, 2021.

[64] V. Puri and N. Kumar, “Wind energy prediction using artificial neural network in himalayan region,” Model. Earth Syst. Environ., pp. 1–10, 2021, doi: 10.1007/s40808-020-01070-8.

[65] P. Jiang, Z. Liu, X. Niu, and L. Zhang, “A combined prediction system based on statistical method, artificial neural networks, and deep learning methods for short-term wind speed forecasting,” Energy, vol. 217, p. 119361, 2021.

[66] W. Li, X. Jia, X. Li, Y. Wang, and J. Lee, “A Markov model for short term wind speed prediction by integrating the wind acceleration information,” Renew. Energy, vol. 164, pp. 242–253, 2021.

[67] X. Liao, Z. Liu, and W. Deng, “Short-term wind speed multistep combined prediction model based on two-stage decomposition and LSTM,” Wind Energy, 2021.

[68] A. T. Peiris, J. Jayasinghe, and U. Rathnayake, “Prediction wind power generation using artificial neural network: ‘Pawan danawi’-A case study from Sri Lanka,” J. Electr. Comput. Eng., vol. 2021, 2021, doi: 10.1155/2021/5577547.

[69] M. Poncela-Blanco and P. Poncela, “Improving Wind Power Forecasts: Combination through Multivariate Dimension Reduction Techniques,” Energies, vol. 14, no. 5, p. 1446, 2021.

[70] J. Maldonado-Correa, M. Valdiviezo-Condolo, M. S. Viñan-Ludeña, C. Samaniego-Ojeda, and M. Rojas-Moncayo, “Wind power prediction for the Villonaco wind farm,” Wind Eng., vol. 45, no. 5, pp. 1145–1159, 2021, doi: 10.1177/0309524X20968817.

[71] S. Shokrzadeh, M. Jafari Jozani, and E. Bibeau, “Wind turbine power curve modeling using advanced parametric and nonparametric methods,” IEEE Trans. Sustain. Energy, vol. 5, no. 4, pp. 1262–1269, 2014, doi: 10.1109/TSTE.2014.2345059.

[72] U. Sharma, “A fine windy day,” Kaggle Dataset, 2021, doi: www.kaggle.com/code/ujjawalsharma20/a-fine-windy-day.

[73] J. Zhou et al., “SDWPF: A Dataset for Spatial Dynamic Wind Power Prediction Challenge at KDD Cup 2022,” arXiv Prepr. arXiv2208.04360, 2022, [Online].

Available: http://arxiv.org/abs/2208.04360.

[74] G. Goretti and A. Duffy, “Evaluation of Wind Energy Forecasts: the Undervalued Importance of Data Preparation,” in 2018 15th International Conference on the European Energy Market (EEM), Jun. 2018, vol. 2018-June, pp. 1–5, doi: 10.1109/EEM.2018.8469845.

[75] W. Yang, J. Wang, H. Lu, T. Niu, and P. Du, “Hybrid wind energy prediction and analysis system based on divide and conquer scheme: A case study in China,” J. Clean. Prod., vol. 222, pp. 942–959, Jun. 2019, doi: 10.1016/j.jclepro.2019.03.036.

[76] J. Maldonado-Correa, M. Valdiviezo-Condolo, M. S. Viñan-Ludeña, C. Samaniego-Ojeda, and M. Rojas-Moncayo, “Wind power prediction for the Villonaco wind farm,” Wind Eng., vol. 45, no. 5, pp. 1145–1159, doi: 10.1177/0309524X20968817, Oct. 2021.

[77] I. Delgado and M. Fahim, “Wind Turbine Data Analysis and LSTM-Based Prediction in SCADA System,” Energies, vol. 14, no. 1, p. 125,, doi: 10.3390/en14010125, Dec. 2020.

[78] Y. Y. Song and Y. Lu, “Decision tree methods: applications for classification and prediction ,” Shanghai Arch. Psychiatry, vol. 27, no. 2, pp. 130–135, 2015, doi: 10.11919/j.issn.1002-0829.215044.

[79] A. Cutler, D. R. Cutler, and J. R. Stevens, “Random forests,” Ensemble Mach. Learn. Methods Appl., pp. 157–175, 2012, doi: 10.1007/9781441993267_5.

[80] A. Sharaff and H. Gupta, “Extra-Tree Classifier with Metaheuristics Approach for Email Classification,” in Advances in Intelligent Systems and Computing, 2019, vol. 924, pp. 189–197, doi: 10.1007/978-981-13-6861-5_17.

[81] M. C. Bottino et al., “Recent advances in the development of GTR/GBR membranes for periodontal regeneration - A materials perspective,” Dent. Mater., vol. 28, no. 7, pp. 703–721, 2012, doi: 10.1016/j.dental.2012.04.022.

[82] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, vol. 13-17-Augu, pp. 785–794, doi: 10.1145/2939672.2939785.

[83] F. Alzamzami, M. Hoda, and A. El Saddik, “Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation,” IEEE

Access, vol. 8, pp. 101840–101858, 2020, doi: 10.1109/ACCESS.2020.2997330.

[84] J. T. Hancock and T. M. Khoshgoftaar, “CatBoost for big data: an interdisciplinary review,” J. Big Data, vol. 7, no. 1, pp. 1–45, 2020, doi: 10.1186/s40537-020-00369-8.

[85] R. E. Schapire, “Explaining adaboost,” in Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik, Springer, 2013, pp. 37–52.

[86] P. K. Chaurasiya, S. Ahmed, and V. Warudkar, “Study of different parameters estimation methods of Weibull distribution to determine wind power density using ground based Doppler SODAR instrument,” Alexandria Eng. J., vol. 57, no. 4, pp. 2299–2311, 2018, doi: 10.1016/j.aej.2017.08.008.

[87] S. M. Malakouti, “Improving the prediction of wind speed and power production of SCADA system with ensemble method and 10-fold cross-validation,” Case Stud. Chem. Environ. Eng., vol. 8, p. 100351, 2023, doi: 10.1016/j.cscee.2023.100351.

[88] T. Gunhan, V. Demir, E. Hancioglu, and A. Hepbasli, “Mathematical modelling of drying of bay leaves,” Energy Convers. Manag., vol. 46, no. 11–12, pp. 1667–1679, 2005, doi: 10.1016/j.enconman.2004.10.001.

[89] M. Verma and H. K. Ghritlahre, “Prediction of Wind Speed by Using Three Different Techniques of Prediction Models,” Ann. Data Sci., pp. 1–33, May 2021, doi: 10.1007/s40745-021-00333-0.

[90] V. Perumpalot, G. V. Drisya, and K. S. Kumar, “Cross-location wind speed prediction for wind energy applications using machine learning based models,” arXiv Prepr. arXiv1808.03480, Aug. 2018, [Online]. Available: http://arxiv.org/abs/1808.03480.

[91] D. Gupta, V. Kumar, I. Ayus, M. Vasudevan, and N. Natarajan, “Short-Term Prediction of Wind Power Density Using Convolutional LSTM Network,” FME Trans., vol. 49, no. 3, pp. 653–663, 2021, doi: 10.5937/fme2103653G.

[92] C.-D. Dumitru and A. Gligor, “Wind Energy Forecasting: A Comparative Study Between a Stochastic Model (ARIMA) and a Model Based on Neural Network (FFANN),” Procedia Manuf., vol. 32, pp. 410–417, 2019, doi: 10.1016/j.promfg.2019.02.234.