الجمهورية الجزائرية الديمقراطيــة الشعبيــة

People's Democratic Republic of Algeria

وزارة التـعليــم العالــي والبحــث العلمــــي

Ministry of Higher Education and Scientific Research

جامعة عين تموشنت بلحاج بوشعيب

University –Ain Temouchent- Belhadj Bouchaib

Faculty of Science and Technology

Department Mechanical Engineering

End of Studies Project

To obtain the Master's degree in : Mechanical Engineering - Energy

Field : Science and Technology

Sector : Mechanical Engineering

Specialty : Energy

Theme

Wind power prediction using Machine Learning models: A comparative study

Presented By:

1) Mr Amar Bensaber Houssam
2) Mr Haddou Benderbal Mohammed Amjed

Before the jury composed of:

Dr REMLAOUI Ahmed		UAT.B.B (Ain Temouchent)	President
Dr OUARI Mohamed Amine		UAT.B.B (Ain Temouchent)	Examiner
Dr Dorbane Abdelhakim	M C A	UAT.B.B (Ain Temouchent)	Supervisor

Academic Year 2024/2025

Dedication

This master dissertation is dedicated to my loving parents and family members whose ongoing love, encouragement, and motivation have been the source of all my efforts.
To my brother and sister, your presence and support have always motivated me to perform at my best.

My heartfelt appreciation to Dr. Abdelhakim Dorbane, supervisor, for his dedication, guidance, and patience in this work.

A special mention of appreciation to everyone in the Mechanical Engineering Department's professors and staff for their passion to teach and for imparting the knowledge and background that made this work possible.

To my dear friends, your companionship and motivation made this experience truly memorable.

Finally, this work is also dedicated to the entire Mechanical Engineering – Energetics promotion, with whom I’ve shared this important academic chapter.

Dedication

This master dissertation is dedicated to my loving parents and family members whose ongoing love, encouragement, and motivation have been the source of all my efforts.

To my brothers , your presence and support have always motivated me to perform at my best.

My heartfelt appreciation to Dr. Abdelhakim Dorbane, supervisor, for his dedication, guidance, and patience in this work.

A special mention of appreciation to everyone in the Mechanical Engineering Department's professors and staff for their passion to teach and for imparting the knowledge and background that made this work possible.

To my dear friends, your companionship and motivation made this experience truly memorable.

Finally, this work is also dedicated to the entire Mechanical Engineering – Energetics promotion, with whom I've shared this important academic chapter.

Acknowledgments

Above all, I would like to thank Almighty Allah for granting me the strength, determination, and perseverance to carry out this thesis successfully.

I would like to acknowledge my sincere thanks to Prof. Abdelhakim Dorbane, Professor of Energetic Mechanics at the Ain Temouchent University. His presence, his patience, and most importantly, his insightful comments were the key factors that helped shape my thoughts throughout this experience.

Sincere thanks also go to the members of the jury who graciously accepted to review my work. Your technical expertise and your time are appreciated.

I also appreciate the entire teaching staff of the University of Ain Temouchent for their helpful guidance and experience.

I particularly thank my lovely parents, whose encouragement has been my pillar of strength throughout my academic life.

Finally, I would like to wholeheartedly express my gratitude to each individual who has directly or indirectly helped in the completion of this work. Your help has been of great significance.

Abbreviations

Arma	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
ARFIMA	Autoregressive Fractionally Integrated Moving Average
PV	Photovoltaic
CSP	Concentrated Solar Power
HAWT	Horizontal Axis Wind Turbine
VAWT	Vertical Axis Wind Turbine
FSI	Fluid-Solid Interaction
LSTM	Long Short-Term Memory
RT	Regression Tree
RF	Random Forest
SVR	Support Vector Regression
MAPE	Mean Absolute Percentage Error
PCA	Principal Component Analysis
FFNN	Feed-Forward Neural Network
MLP	Multi-Layer Perceptron
GW	Gigawatt
MW	Megawatt
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
ANN / ANNs	Artificial Neural Network(s)
SVM / SVMs	Support Vector Machine(s)
SCADA	Supervisory Control and Data Acquisition

SSA	Singular Spectrum Analysis
SHAP	SHapley Additive exPlanations
DBN	Deep Belief Network
HLNN	Hybrid Laguerre Neural Network
XGBoost	eXtreme Gradient Boosting
CatBoost	Categorical Boosting
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
MSE	Mean Squared Error
R^2 or R2	Coefficient of Determination
IQR	Interquartile Range
AR	AutoRegressive

Abstract

This study addresses the challenge of integrating wind energy into power systems due to its natural variability by employing machine learning techniques to predict wind power generation using real field data from a Senvion MM82 wind turbine in France. Five algorithms—Random Forest, Decision Tree, XGBoost, CatBoost, and Linear Regression—were evaluated using standard performance metrics, with Linear Regression emerging as the most accurate. The research highlights the critical role of data preprocessing, feature engineering, and the inclusion of meteorological and operational parameters, ultimately demonstrating the potential of machine learning to enhance forecasting accuracy, support grid integration, and improve the reliability of renewable energy systems.

Keywords: Wind energy, machine learning, artificial intelligence, forecasting

Résumé

Ce mémoire fin d'études aborde le défi de l'intégration de l'énergie éolienne dans les réseaux électriques en raison de sa variabilité naturelle, en utilisant des techniques d'apprentissage automatique pour prédire la production d'énergie éolienne à partir de données réelles collectées sur une éolienne Senvion MM82 en France. Cinq algorithmes — Random Forest, Arbre de Décision, XGBoost, CatBoost et Régression Linéaire — ont été évalués selon des métriques de performance standard, la Régression Linéaire s'étant révélée la plus précise. L'étude souligne l'importance du prétraitement des données, de l'ingénierie des caractéristiques, ainsi que de l'intégration des paramètres météorologiques et opérationnels, démontrant ainsi le potentiel de l'apprentissage automatique pour améliorer la précision des prévisions, faciliter l'intégration au réseau et renforcer la fiabilité des systèmes d'énergie renouvelable.

Mots clés : Énergie éolienne, apprentissage automatique, intelligence artificielle, prévision

ملخص

تتناول هذه الأطروحة في مرحلة الماجستير تحدي دمج طاقة الرياح في شبكات الكهرباء بسبب طبيعتها المتقلبة، من خلال استخدام تقنيات التعلم الآلي للتنبؤ بإنتاج طاقة الرياح اعتمادًا على بيانات ميدانية حقيقية تم جمعها من توربين رياح من نوع Senvion MM82في فرنسا. تم تقييم خمسة خوارزميات — الغابة العشوائية، شجرة القرار، XGBoost ، CatBoost، والانحدار الخطي — باستخدام مقاييس أداء معيارية، وقد تبيّن أن الانحدار الخطي هو الأكثر دقة. وتُبرز الأطروحة أهمية معالجة البيانات المسبقة، وهندسة الميزات، وإدماج المعلمات المناخية والتشغيلية، مما يُظهر قدرة التعلم الآلي على تحسين دقة التنبؤات، ودعم دمج الطاقة في الشبكة، وتعزيز موثوقية أنظمة الطاقة المتجددة.

الكلمات المفتاحية: طاقة الرياح، التعلم الآلي، الذكاء الاصطناعي، التنبؤ.

Table of contents

List of figures

List of tables

General Introduction

Renewable energy technology has become crucial for addressing climate change and energy supply challenges. Since the 1980s, wind energy has significantly contributed to renewable energy sources **[1]** ,It plays a crucial role in satisfying the rising global demand for electricity and is an important sector of the global renewable energy industry. As we see in the Figure 1, the four main types of renewable energy are Wind energy, Solar energy , Ocean energy , Biomass energy.

Figure 1:The four main renewable energy

The transition to renewable energy sources has become an unavoidable factor due to the degradation of environmental conditions and exhaustion of fossil fuel resources. Renewable energy from resources like solar, wind, biomass, and ocean energy offers sustainable alternatives; wind energy is an important contributor because it is scalable and technologically mature[2]

All mega-projects in the world have been rapidly implemented due to their comparative advantages concerning wind power against carbon emission reductions, operational costs, and energy security-for example, Gansu Wind Farm 7,965 MW, Hornsea Offshore Wind Farm 1,800 MW[3]

Wind power capacity has shown a rapid increase recently and has become a promising source of renewable energies. For example, 8.4% of the total U.S. utility-scale electricity generation is provided by wind turbines in 2020, and it is expected to reach 20% by 2030 and 35% by 2050 [4]. The significant advantage of wind energy is avoiding approximately 189 million metric tons of CO2 emissions and reducing water usage by about 103 billion gallons compared to traditional power sources [4].Figure 2 shows the general layout of an onshore wind farm .

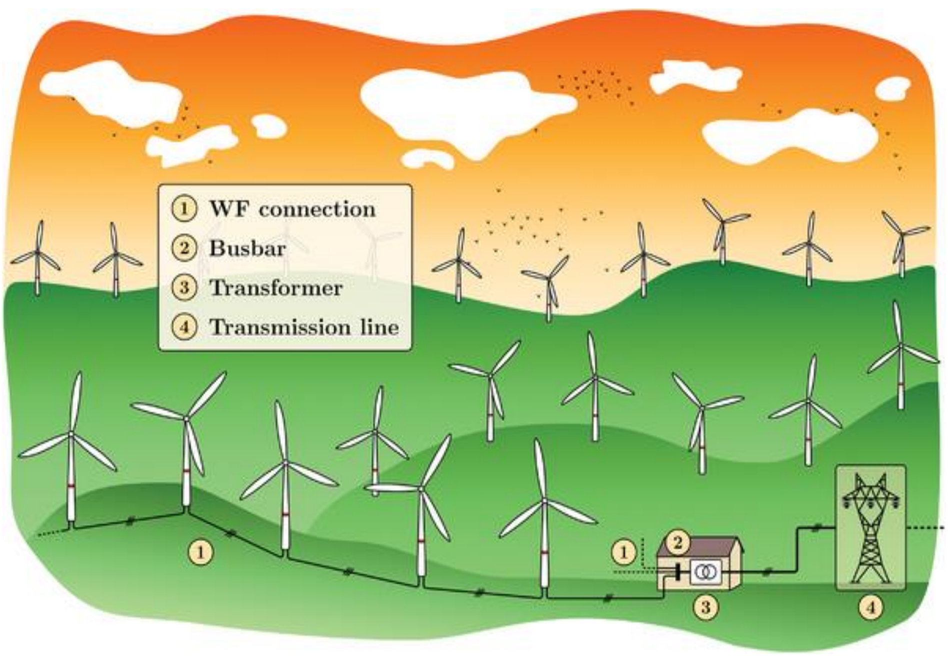


Figure 2:the general layout of an onshore wind farm[5]

However, the main challenges in wind power management reside in its intermittent fluctuations mainly due to weather conditions, which makes its integration into a power grid a challenging task [6]. Thus, predicting wind power is undoubtedly necessary for the efficient integration of wind turbines into the power grid. Over the last two decades, there has been increasing interest in developing accurate wind power prediction methods[7], Two main types of models can be distinguished: physical-based and data-driven models [8], Physical models employed atmospheric motion equations for estimating the evolution of meteorological measurements, and then they used the estimated meteorological variables for wind power prediction[9], Essentially, predicting wind power via a physical model using numerical weather estimation is accomplished in two phases: at first, wind speed is predicted and then transformed into wind power[10], However, physical models are generally costly and time-consuming to design, producing low prediction precision for a local area [11].

In contrast to physical approaches based on very complex differential equations, the data-based models derive functional dependencies directly from the data to build a model that describes the relations between wind power and other input variables [12]

Efficiently predicting wind power is vital to help operators integrate wind turbines in smart-grids and improve the management of power output. Our study attempts to predict wind power generation from Senvion MM82 dataset using five common machine learning models, namely : Random forest , Decision Tree , XGBoost , CatBoost , Linear Regression. The remaining manuscript is organized as follows; chapter 2 gives a bibliographic study on the prediction of wind energy using machine learning , chapter 3 describes our research methodology , including the input and the output parameters and and their importance in predicting wind power, and the selection of machine learning models , chapter 4 focuses on the simulation part including evaluation metrics results and discussion. Our manuscript is included by conclusions and perspectives.

Chapter I. Generalities

This chapter explores the growing role of renewable energy in contributing to a solution for the environment and sustainability issue caused by the increasing consumption of fossil fuels. As global energy demands rise with population growth and technological advancement, traditional sources of energy—primarily fossil fuels—are making a huge contribution to climate change and global warming. In response to this, researchers have proposed several measures to mitigate their effects, including improving energy efficiency, developing clean energy technologies, and alternative fuel utilization.

Among these alternatives, renewable energy is the most viable and promising long-term solution. This chapter introduces the most significant sources of renewable energy—solar, wind, biomass, and ocean energy—with their technologies, real-world applications, and prominent worldwide projects. Wind power is treated in greater detail, including its mechanisms, turbine technologies, large-scale implementations, as well as wind power output forecasting using machine learning.

By giving an overview of all these renewable energy technologies and resources, this chapter indicates how they have the potential to substantially substitute for fossil fuels and achieve a more sustainable energy future.

I.1 Renewable energy

Increasing population and technological development have resulted in high consumption of fossil fuels that are non-renewable resources, and these pose severe environmental threats, especially to global warming and climatic changes [13]. This situation has aroused researchers all over the world to look for means of reducing or totally eliminating the use of fossil fuels. In order to reduce dependence on fossil fuels and their contribution to climate change, there are three effective ways that may help, namely: (a) efficiency improvement of conventional power generation systems by means of waste heat recovery [14]; (b) development of environmentally friendly energy conversion systems, including fuel cells [15]; and (c) alternative fuels from natural resources with less harmful environmental impact [16]. Of the alternatives, renewable energy is the most promising avenue to accept and can probably enable the reduction or complete elimination of dependency on fossil fuels.

In the last decade, phenomenal growth has been observed in the development and deployment of renewable energy technologies. Various renewable sources such as solar [17], wind [18], biomass [19], and ocean energy [20] are already in use or under significant development for large-scale power supply. Some of these major projects that have been executed across the world are presented in Table 1.

Table 1:Examples of the largest renewable energy projects of different technologies around the world:

Power Plant Name	Technology	Country	Year	Installed Capacity (MW)	Reference
Three Gorges Dam	Hydroelectric Power	China	2003	22,500	[21]
Itaipu Dam	Hydroelectric Power	Brazil and Paraguay	1984	14,000	[22]

Bhadla Solar Park	Photovoltaics	India	2018	2245	[23]
Longyangxia Dam Solar Park	Photovoltaics & Hydroelectric Power	China	2015	2130	[24]
Huanghe Hydropower Hainan Solar Park	Photovoltaics	China	2020	2200	[25]
Gansu Wind Farm	On-ShoreWind Farm	China	2009	7965	[26]
Alta Wind Energy Center	On-ShoreWind Farm	United States	2010	1550	[27]
Muppandal wind farm	On-ShoreWind Farm	India	_	1500	[27]
Ironbridge power plant	Biomass Power Plant	United Kingdom	2012	740	[28]
Alholmens Kraft Power Plant	Biomass Power Plant	Finland	2002	240	[29]
Polaniec biomass power plant	Biomass Power Plant	Poland	2012	220	[30]
Ouarzazate Solar Power Station	Parabolic trough and solar power tower (CSP)	Morocco	2016	580	[31]

Ivanpah Solar Power Facility	solar power tower (CSP)	United States	2014	377	[32]
Mojave Solar Project	Parabolic trough (CSP)	United States	2014	280	[33]

I.1.1 Types of renewable Energy

I.1.1.1 Solar Energy

The world can boast of solar energy as the most accessible source of renewable energy. It covers all types: solar thermal [34] and solar photovoltaic (PV) technologies [35], which have already been implemented in so many areas, such as use on houses [36], desalination [37], transportation [38], drying [39], and irrigation[40]. Various research projects have sought to elaborate on these developments and their practical applications. Some of the most promising options that would deliver both heat and electricity are the Concentrated Solar Power technologies. Major investment was poured into developing CSP worldwide with tens of commercial systems entering service since 2005. Much remains to be done in order to make the CSP a reliable and affordable source of energy. There are two main types of solar collectors: line focus, which generates solar energy along a collector's focal length, and point focus, which concentrates sunlight onto a point. Examples of line-focus collectors include parabolic troughs and linear Fresnel reflectors. Examples of point-focus collectors include solar thermal towers and parabolic dishes [41]. A schematic of these collectors is shown in **Figure 3**.

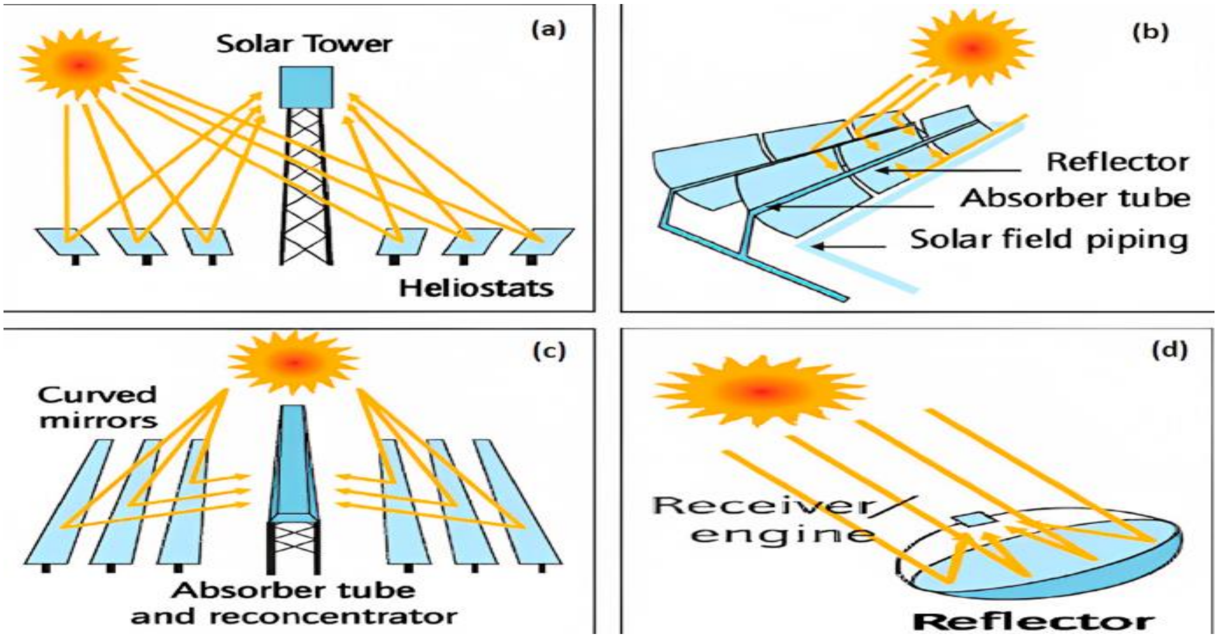


Figure 3: Recent available CSP technologies: (a) solar thermal tower, (b) parabolic trough collectors ,(c) linear fresnel reflectors, (d) parabolic dish [42]**.**

I.1.1.2 **Wind Energy:** We will discuss it in section I.2

I.1.1.3 Biomass Energy

Biomass follows a wide range of applications, starting from traditional approaches in combustion methods [43] to its direct usage in fuel cells for producing electricity [44]. Although biomass is commonly burned, it may not be the most effective method for using the full energy potentiality of various biomass resources. Converting biomass into biofuels, such as biodiesel [45] and biochar [46], offers promising alternatives that can be used over a wide range of applications.
In this Figure 4 we will see different types of biomass energies are available:

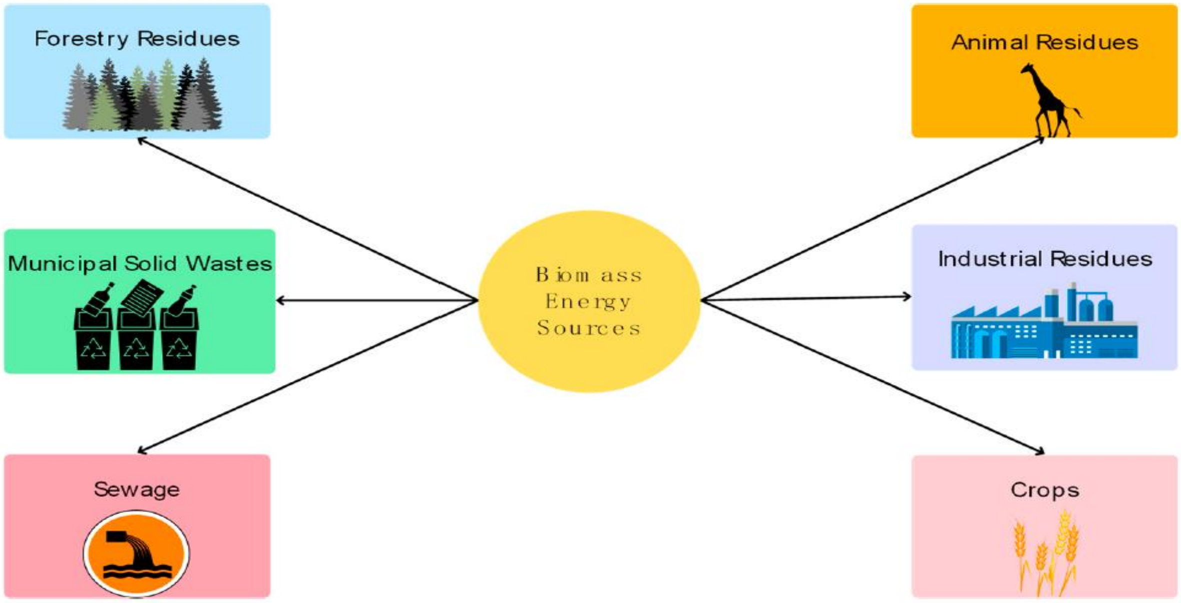


Figure 4: Biomass energy sources.

I.1.1.4 Ocean Energy

Current efforts in renewable energy increasingly focus on harnessing ocean energy, since oceans constitute about three-quarters of the Earth's surface. Indeed, much research is currently in place to unlock the full potential of this promising source of energy [29]. Some of the leading approaches under consideration include wave power and tidal currents, which are very consistent and predictable, thus more suitable for electricity generation than other renewable energy sources [47] . Whereas the theoretical potential of marine energy could cover world energy needs, present ocean energy is developed modestly. Significant research is in progress to find out the most suitable sites to harness the energy resource, quantify the efficiency of ocean energy conversion devices, quantify the power that can be delivered, and study the environmental effects. As the industry is expanding, governments worldwide are implementing policies to encourage the use of ocean energy [48] , like we see in Figure 5 The PacWave's first marine energy project in Oregon.

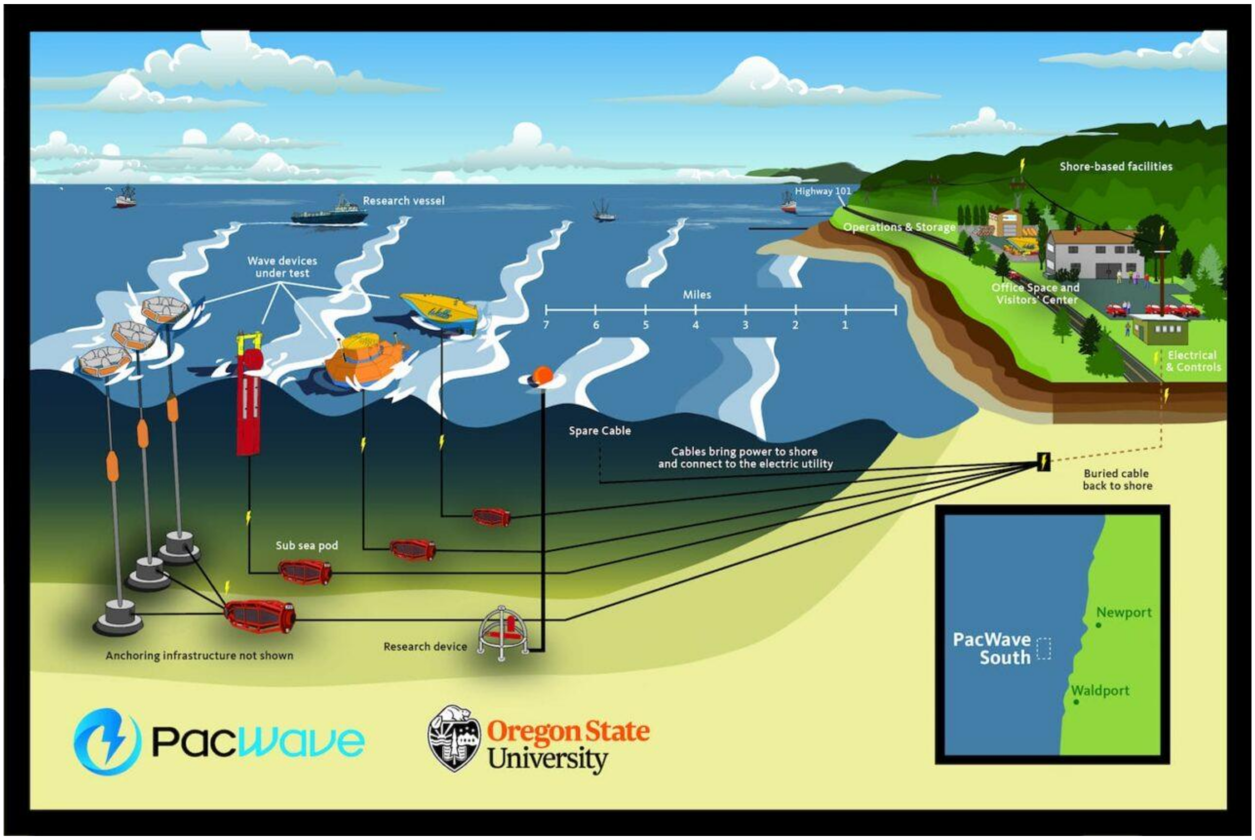


Figure 5: PacWave's first marine energy project in Oregon

I.2 wind energy

I.2.1 Concept of Wind Energy

Wind energy refers to the kinetic energy harnessed from air movement in the atmosphere, wherein wind turbines are used to transform this kinetic energy into electrical energy (Boyle, 2012). [49]

I.2.2 Understanding Wind Turbines: Components, Functionality, and Types

The main components of a wind turbine are its tower, blades, and the nacelle, which houses the generator, gears, and control system. Similar to how an airplane wing provides lift for a plane, the wind propels the blades into motion. The generator inside the nacelle receives energy from the turbine through the driving shaft. After the generator has converted the kinetic energy that is being produced into electrical energy, the transformer will deliver that energy to the grid,As can be seen in **Figure 6** , the two primary classifications of present wind turbines are known as horizontal axis wind turbines (HAWT) and vertical axis wind turbines (VAWT). The HAWTs dominate the majority of the wind sector since they are more efficient and produce more electricity than VAWTs. Since VAWTs are situated close to the ground and are consequently less exposed to the wind, reducing power production, they are inherently unreliable [50].

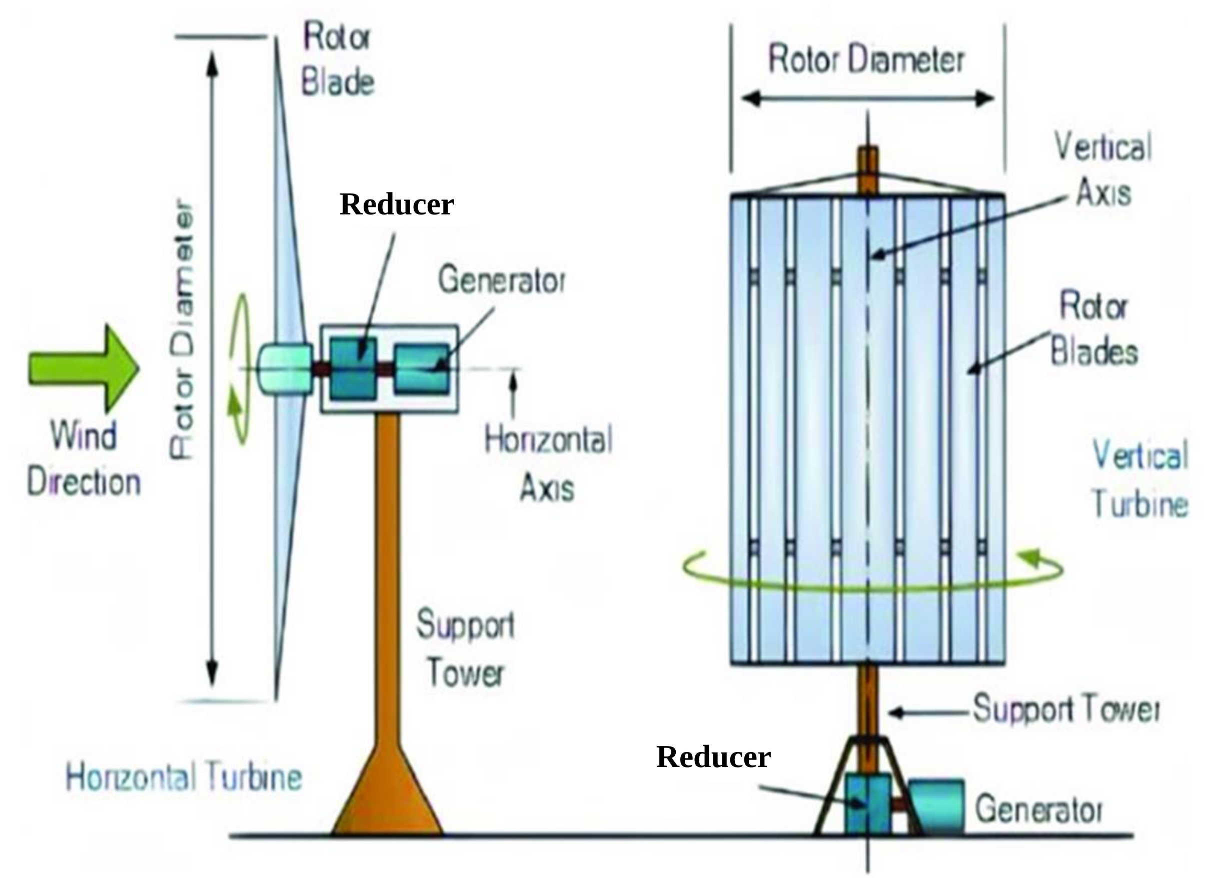


Figure 6: Wind turbine types: VAWT and HAWT [51]**.**

By converting the wind's kinetic energy, wind energy may either be transformed directly into mechanical power or indirectly into electrical energy. The wind turbine is an essential part of any wind energy system since it is the component responsible for converting the potential energy of the wind into a form of mechanical power that can then be used in various contexts. At the beginning of the 20th century, construction began on the first wind turbine designed to generate electrical power. Despite the fact that wind turbine technology has been steadily advancing, enormous advances have been particularly achieved in wind turbine design[52].

Modern technical advancements and refinements of a turbine and its components have resulted in considerable improvements in produced power production and efficiency. Furthermore, advancements in particular generators, as well as the usage of power electronic devices, have enabled gearless turbine designs[53].

Several large-scale wind projects based on two technologies, i.e., on-shore and off-shore, implemented worldwide are shown in Table 2

Table 2**:** Largest wind energy projects around the globe.

Power plan name	Technology	country	year	Installed capacity(MW)	reference
Gansu Wind Farm	On-Shore Wind Farm	China	2009	7965	[3],[54]
Gansu Wind Energy Center	On-Shore Wind Farm	United States	2010	1550	[54]
Muppandal Wind Farm	On-Shore Wind Farm	India	1986	1500	[54]
Hornsea Project Two	Off-Shore Wind Farm	United Kingdom	2022	1800	[55]
Hornsea Project one	Off-Shore Wind Farm	United Kingdom	2019	1200	[55]

The Jiuquan Wind Power Base, also known as the Gansu Wind Farm Project, is a group of huge wind farms currently being constructed on the western side of the Gansu province in China. The Gansu Wind Farm Project is located in a desert area adjacent to the city of Jiuquan at two distinct places in Guazhou County, as well as close to Yumen City, in the northwest province of Gansu (which has an availability of wind power). These sites are part of Guazhou. The complex only uses 7.965 MW of its installed capacity, while it has a projected capacity of 20 GW and is now running at less than 40% capacity utilization of its existing 8 GW [3].

The Alta Wind Energy Center in California has been crowned the wind farm with total turbines of 1320 MW, and there are plans to extend it to 3000 MW in the future. At the end of 2012, it had 440 wind turbines as part of its infrastructure. In the United States, 815 wind farms have been placed into service, with a total power capacity of 60 GW. This is sufficient to power 15 million residences in the United States. Wind power has overtaken solar as the leading source of newly installed capacity in the United States. The United States has reclaimed its position as the world's leading wind power market. China's wind power capacity has hit 75 GW, making it the nation with the highest installed capacity globally [56].

The Hornsea Project Two wind farm, which will become one of the largest offshore wind projects, received approval from the British government in August 2016, allowing it to continue with its construction. It can generate 1.8 GW of electricity when fully operational. This project is an expansion of the 1.2 GW Hornsea Project (offshore wind farm), which was already cleared for construction. The Hornsea Project comprises 300 wind turbines made by Dong Energy, is now under development and will produce enough energy to power 1.6 million households. The project's location will be around 90 kilometers (55 miles) east of Yorkshire's coast in England [55].

I.2.3 Advantages of Wind Energy

Renewable source of energy: It is inexhaustible and available in many parts of the world (IPCC, 2011) [57]

Carbon emission reduction: Wind energy is clean since it does not emit harmful emissions to the atmosphere, hence environmental-friendly (Sovacool, 2009) [58].

Low operating cost: It requires very minimal maintenance costs after installation compared to conventional sources of energy (GWEC, 2020) [59].

I.2.4 Disadvantages and Challenges of Wind Energy

Wind speed variability: This leads to fluctuations in energy production (Ackermann & Söder, 2002) [60].

Impact on wildlife: Wind turbines kill birds and bats (Kuvlesky et al., 2007) [61].

High initial costs: It is pretty capital-intensive to establish wind farms (Wiser & Bolinger, 2019) [62].

I.2.5 Future Developments in Wind Energy

Wind energy technology is constantly evolving. Among the progresses made in wind energy is the invention of floating wind turbines that will be installed in deep seas and improved storage technologies for overcoming fluctuating energy generation (IRENA, 2021) [63].

I.3 The role of machine learning in predicting wind power

Machine learning provides a fresh perspective on the prediction of energy generation and performance of wind power installations, thus acting as a viable alternative to conventional analytical models. Most conventional methods rely on complex systems of differential equations [64] .that require heavy computational resources and generally take longer to yield results with acceptable accuracy. By embracing machine learning, it's possible to discern patterns within multi-dimensional data in such a way that the resulting models are robust, resistant to outliers, and handle noisy data well.

To solve this, techniques to be used include SVR, RT, RF, and ANNs. The methods of prediction will differ depending on the strategy to be followed-physical, statistical, machine learning, or a hybrid that incorporates elements of these approaches. The predictions may also differ, being either direct, where the output power is forecasted directly, or indirect, where wind direction and speed are predicted and output power derived from power curves. Other ways to distinguish between prediction techniques are by the time frame over which they predict: short term, medium-term, and long-term forecasts. Physical modeling deals with complicated environmental attributes around the wind turbines, hence, involving the identification, collection, and pre-processing of those physical characteristics. It enjoys a lesser demand on historical data compared to statistical models. However, such systems are kind of complex while developing and operating [65]).

In the view of statistical modeling, it becomes necessary that prior to modeling proper, there exist certain assumptions to be satisfied which, sometimes conflict with the reality of the nature of wind-a non-stationary phenomenon. ANN, regression trees, Random Forests, and Support Vector Regression can equally be applied in nonlinear modeling jobs such as prediction of wind energy. Other models are directed toward minimizing the error across the training set, while SVR is toward fitting as many data points as possible within a specified error margin, hence focusing on minimizing the upper bound of expected risk through structural risk minimization. Meanwhile, RT, RF, and ANN depend on empirical risk minimization.

A pragmatic approach to establishing error intervals based on general noise functions can be developed, allowing us to select model parameters effectively and refine them for data fitting [66].It's important to note that SVRs, like RT and RF, are sensitive to the chosen input variables, necessitating thorough data preparation before feeding them into the models. Techniques like wavelet transformation help break down time series data into components that are more stationary [67]. An efficient variant of SVR, known as least-squares SVR, simplifies the problem to a linear one, which makes it more computationally efficient than the standard SVR or ANN.
Coupled with wavelet transformation, it has shown the ability to outperform ANN for wind energy forecast as far as 24 hours ahead. Conventional SVR method can also be used by using a Euclidean distance-based approach to choose and identify the data segments whose patterns are most closely related to the reference forecasting sequence [68]. Chosen segments can then be used for model training. Unlike some approaches, an indirect method is adopted here, treating wind speed as the dependent variable. By predicting wind speed and applying power curves, we can estimate the actual wind energy generated. It's important to recognize that fluctuations in wind energy production, such as sudden ramps, can negatively impact the grid. Therefore, timely predictions are essential for implementing optimal strategies to mitigate adverse effects on grid stability [69].

In order to enhance the prediction performance of ramps using SVR, the importance of each variable is evaluated and the selection process is simplified. An orthogonal test is adopted, which has shown better results compared to other methods like Spearman correlation, Gray correlation, and principal components analysis. All four methods have been applied with the SVR model. Ensemble techniques, such as Random Forest, have promise for accurate wind energy forecasting and easily conduct feature importance assessments [70]. As a non-parametric model, RF does not require extensive parameter tuning, often a resource-intensive process. A further refinement, Quantile RF, extends standard RF and produces confidence intervals around its predictions [71]. It is also underlined that the selection of appropriate input data is necessary, as well as the issue must be treated from the viewpoints of correlation and importance of the features. We are thus enabled to model just those features bearing valuable information concerning the problem at hand [72]. ANNs represent a very useful tool in many applications, such as pattern recognition, forecasting, monitoring, and optimization of control and design in the sector of wind energy.

For effective prediction, one needs to ensure that the independent variables capture the underlying intricacies of the phenomenon being studied. At face value, it may seem as though the more the variables, the better the outcome; however, this poses a different set of problems which, if not managed properly, will lead to several disadvantages resulting from too many variables. As the number of variables increases, so does the data needed to properly analyze the data increase correspondingly [73] ,This increase in information volume tends to have adverse generalization, besides overfitting or underfitting issues with increased requirements in training time and computational resources.

Principal Component Analysis, or PCA, is really helpful when several variables need to be handled yet retaining the majority of the information [74]

A number of scientific researches suggest that ANN can be efficiently used in order to cope with complex problems [75] Bursts happen seldom and their prediction is also difficult, thereby affecting the result of forecasting [76] . However, in cases without bursts, ANN has proved its efficiency. Comparing forecasted values with actual wind energy produced after the forecast, maximum error obtained was 6.52%. Neural network used in forecasting was optimized using genetic algorithms and better forecasts were obtained than those without such optimization. Another architecture which is seen to be quite promising for forecasting electric power produced by wind turbines is the multi-layer perceptron (MLP) [77]. This forecast was performed for a short-term horizon using the Levenberg-Marquardt learning algorithm. The training and test sets had a MAPE less than 3.26%. An interesting approach is to collect and use data from two different locations [78] with different characteristics-one site with high winds and the other with low winds-within a single model. Further, another study has used a feed-forward neural network (FFNN) and came to the conclusion that the prediction accuracy depends on the time horizon selected. Specifically, the accuracy of long-term forecasts tends to decline, while that of short-term ones improves [79].

Figure 7: Forecasting Wind Patterns Using Machine Learning [80]

Chapter II. Bibliographic study

This chapter highlights recent research work on wind power forecasting using Artificial Intelligence techniques, i.e., Machine Learning (ML) and Deep Learning (DL). According to the study, models such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), ensemble-based methods, and hybrid methods have been proven to improve short- and medium-term wind power prediction. Key conclusions suggest that the inclusion of meteorological information and SCADA system outputs enhance prediction accuracy by a significant amount. Ensemble and hybrid strategies, especially those involving data filtering or signal decomposition techniques (e.g., wavelet transform, SSA), outperform standard models in dealing with variability and noise in wind data. Further, research underscores the importance of model interpretability, and interpretable models are utilized by leveraging techniques such as SHAP for explaining model choices. Overall, the integration of advanced AI methodologies provides stable, precise, and scalable solutions to wind power prediction, allowing renewable energy systems to be optimized and stabilized.

II.1 Previous studies on wind power prediction using Artificial Intelligence (Machine Learning and Deep Learning)

Various data-driven approaches have been developed in the literature to improve wind power prediction. Traditional time-series methods, including the autoregressive moving average (ARMA) model and its variants, have been widely used for short-term wind power forecasting [81]**,** The method in [82]used an ARMA model to forecast hourly wind power. It showed good forecasting performance for one hour ahead and declines in precision further along in time. These models are simple to construct and convenient to implement. However, it is worth while noticing that traditional time-series models (e.g., ARMA and its variants) can reach a satisfying performance when wind power data show regular variations, but the forecast error is obvious when the wind power time series shows irregular variations. In [83], a coupled strategy integrating ARMA and an artificial neural network (ANN) has been proposed to short-term forecast wind power. This study showed that the coupled approach provided a better forecasting performance compared to the standalone ARMA and ANN.As we see in the **Figure** 8 Block diagram of ARIMA (0,1,1) model of wind power time series.

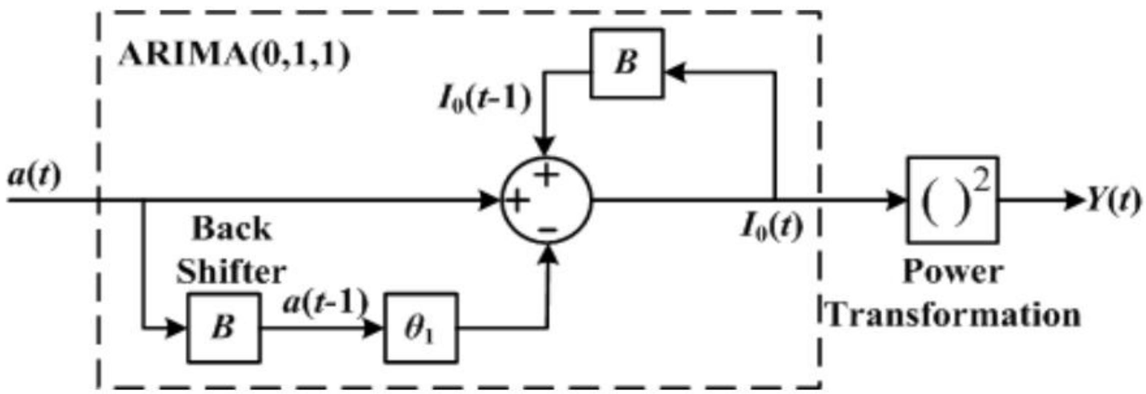


Figure 8: Block diagram of ARIMA(0,1,1) model of wind power time series.[84]

On the other hand, several studies have investigated the wind power prediction using machine learning and deep learning techniques (Artificial Intelligence). For instance, Nielson et al. [85] conducted a research to enhance the accuracy of power prediction from wind turbines using atmospheric information in artificial neural network (ANN) models. Machine learning techniques, ANNs, trained with atmospheric input parameters (e.g., wind speed, temperature, pressure, turbulence) and turbine operation data were employed by the authors. Their strategy demonstrated that using atmospheric conditions greatly enhances prediction accuracy compared to turbine data-based models. The results presented enhanced accuracy of short-term power forecasting, reducing errors and closely approximating wind power generation variability. The study concluded that applying atmospheric inputs into ANNs is a valid option for optimizing the efficiency of wind energy, improving grid management, and ensuring reliability in renewable energy. The technique bridges loopholes in traditional models by addressing the significant role that environmental parameters contribute to energy prediction.

In another study, A. Alkesaiberi et al. [86] conducted a comparative analysis to evaluate the accuracy of machine learning (ML) methods for wind power prediction. The authors contrasted and compared several ML models, i.e., support vector machines (SVM), random forests, and artificial neural networks (ANN), against meteorological inputs and past wind power output. The authors highlighted the importance of feature and hyperparameter optimization in their method for enhanced prediction accuracy. The results showed that models based on ANN performed better than other methods regarding accuracy and reliability, particularly in detecting nonlinear correlations in wind power data. The study concluded that machine learning, specifically ANNs, offers a strong platform for accurate wind power prediction, which can optimize energy output and grid stability. Additionally, the authors emphasized that integrating meteorological parameters into forecast models will maximize the efficiency and scalability of renewable energy systems. Their work recommends the application of advanced ML techniques to support sustainable energy management.

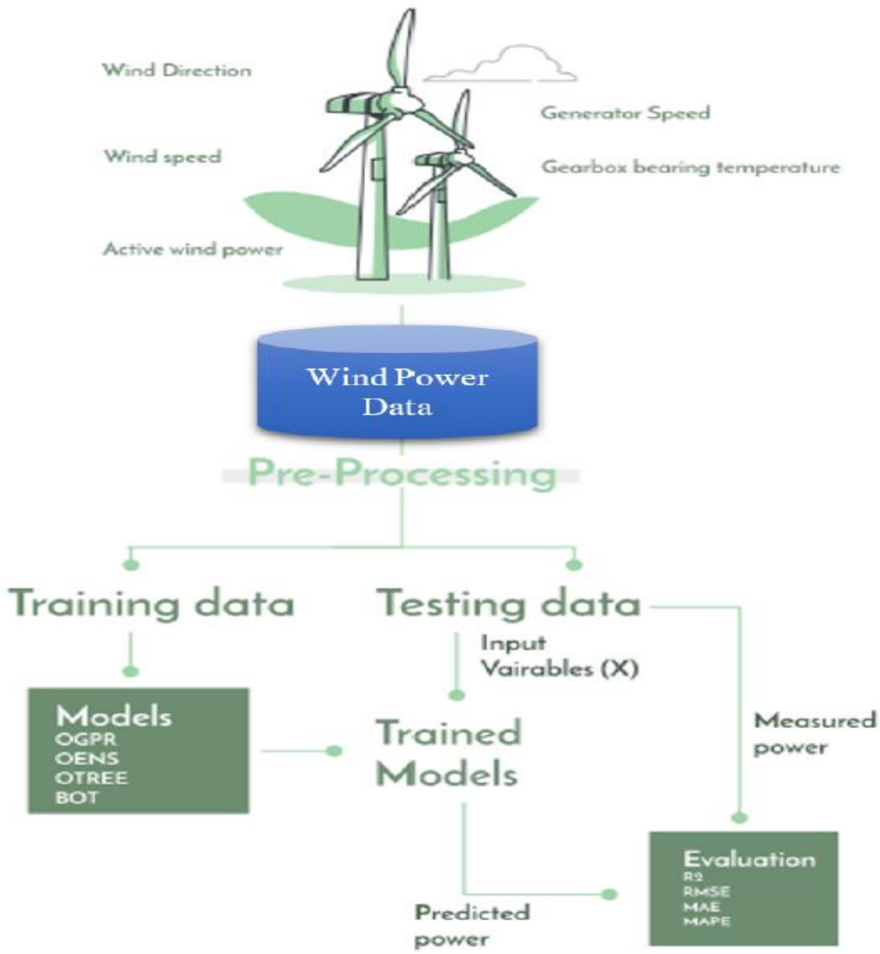


Figure 9 : Machine learning-based wind power forecasting framework [86].

In [87], Singh et al. conducted a work to improve power prediction of wind turbines based on SCADA system data and machine learning (ML) techniques. The authors studied big SCADA data sets in order to establish the most influential operational and environmental factors affecting the performance of a turbine. Regression algorithms and ensemble techniques were utilized by the authors for power prediction and detection of performance anomalies based on ML models. The findings revealed that ML-based techniques, especially ensemble models, demonstrated excellent accuracy in short-term power forecasting and fault detection, superior to conventional statistical models. The research concluded that combining SCADA data with ML improves predictive precision, facilitating proactive maintenance and maximized energy generation. The authors identified the potential application of ML-based SCADA analytics to reduce downtime, extend the life of the turbines, and improve grid stability, advocating its application in wind power management systems.

Furthermore, Lee et al. conducted this work to enhance the prediction of wind power using ensemble learning techniques. The authors proposed and compared various ensemble models, including Random Forest, Gradient Boosting, and Stacking, with meteorological data and historical turbine operation data from SCADA systems. The strategy was centered on combining base learners (e.g., decision trees, neural networks) to enhance the stability of predictions. The results demonstrated that ensemble models, particularly Stacking-based models, were more stable and accurate compared to individual machine learning methods, with reduced error rates in short- and medium-term forecasts. The study concluded that ensemble learning frameworks significantly improve wind power prediction reliability by avoiding model-specific weaknesses and detecting complex data patterns. The authors highlighted the importance of such models in practice to achieve the optimal integration of wind energy into power systems, reduce operating uncertainties, and enhance sustainable energy planning. They also advocated for extensive use of ensemble approaches in renewable energy systems for enhancing forecasting accuracy and grid resilience [88].

Also, Cakiroglu et al. [89] conducted a study to develop interpretable ensemble learning models for wind turbine power prediction with the inclusion of SHAP (SHapley Additive exPlanations) analysis for transparency. The authors used data-driven ensemble methods, namely Random Forest, Gradient Boosting, and Stacking, trained on SCADA-derived operation and meteorological datasets. The authors incorporated SHAP values for feature importance quantification and model decision explanation to render the models more interpretable. The results indicated that ensemble models had better predictive accuracy compared to individual-algorithm approaches, whereas SHAP analysis revealed the most important predictors to be wind speed, turbulence, and turbine operating parameters. The study concluded that the combination of ensemble learning and SHAP not only provides high accuracy but also interpretable insights into the dynamics of wind power, fostering trust in ML-driven predictions. The authors emphasized the necessity of interpretable models for operation optimization, maintenance planning, and grid integration, and advocated for their application in real-world wind energy systems to sacrifice performance for interpretability.

On the other hand, a study was conducted by S. Liu et al. [90] to enhance short-term wind turbine power prediction by integrating machine learning (ML) with digital filtering techniques. The authors employed ML techniques, such as artificial neural networks (ANNs) and support vector machines (SVM), along with digital filters (i.e., low-pass or wavelet filters) for the filtration of noisy SCADA data as well as meteorological inputs. Their solution was to reduce signal noise and extract informative temporal patterns to enhance prediction performance. The result indicated the hybrid framework (ML + digital filter) significantly outcompacted single ML models with enhanced precision and stability in 24-hours-ahead prediction. The article concluded that prefiltering raw data with digital filters is highly effective at eliminating uncertainties due to noise, thereby enhancing ML model reliability in wind power prediction. The authors highlighted the practical benefits of this approach to grid operators through enhanced energy dispatch planning and the scheduling of turbine maintenance. They encouraged the application of hybrid methods in addressing real-world data quality problems in renewable energy systems.

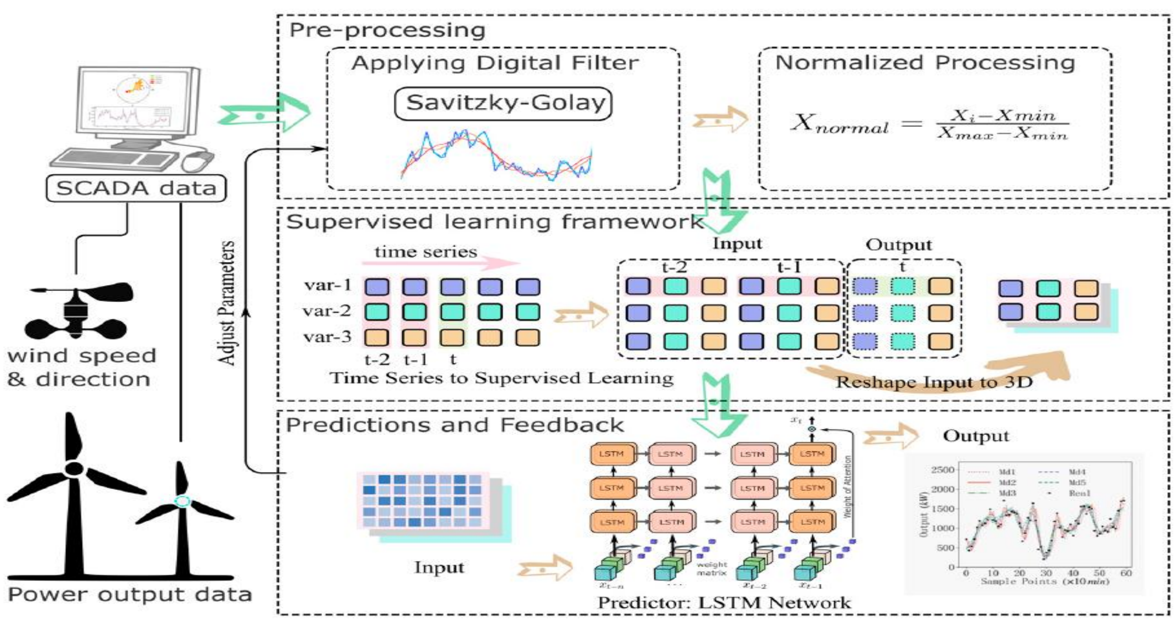


Figure 10: Data flow diagram of wind power short-term forecasting task framework [90]

Moreover Jiajun et al. [91] conducted a study to improve ultra-short-term wind forecasting precision by combining wavelet transform, deep belief networks (DBN), and ensemble learning. They proposed a hybrid model that employs wavelet transform to decompose the wind speed signals to capture multi-scale features, DBN to extract high-level features and predict, and ensemble learning for added robustness. The results demonstrated that their approach outperformed traditional models in predictive accuracy and stability, particularly in handling non-stationary wind data. They concluded that the combination of signal decomposition, deep learning, and ensemble techniques is appropriate for reliable wind energy prediction, which can simplify grid operation and the integration of renewable energy.

Additionally Wang et al. [92] conducted a study to enhance the precision of wind power forecasting by creating a hybrid model that integrates singular spectrum analysis (SSA) and a hybrid Laguerre neural network (HLNN). SSA was utilized to break down raw wind power data into trend, periodic, and noise components in order to reduce non-stationarity. HLNN, which involves integrating Laguerre filter structures into neural networks, was subsequently employed to model complex temporal patterns in the decomposed data. Their results indicated that the hybrid model possessed higher prediction accuracy and stability compared to conventional methods, with lower mean absolute error and improved robustness for short-term forecasting. They concluded that the coupling of SSA and HLNN is an efficient method for resolving issues posed by noisy and irregular wind power data and is a trustworthy tool for optimizing the grid integration of renewable energy.

As well Harrou et al. [93] conducted this study using data-driven technologies to monitor solar and wind power systems. They applied modeling and data analysis to improve anomaly detection and performance optimization of the systems. The results obtained show a significant improvement in monitoring accuracy and reduced downtime. The authors conclude that such data-driven technologies are efficient in making renewable energy more reliable and efficient, thereby facilitating a more sustainable energy transition.

In [94] Hanifi et al. offered a critical overview of wind power forecasting methods, their historical evolution, current trends, and future trends. They elaborated on various methods, e.g., statistical methods, physical models, artificial intelligence methods, and hybrid methods. The authors concluded that machine learning-based methods and hybrid methods are more accurate compared to traditional methods. They also stressed the importance of including meteorological data and wind farm parameters to improve forecasts. Lastly, Hanifi et al. highlighted the current issues of wind volatility and model complexity and suggested that future research should be directed towards improving the robustness, scalability, and interpretability of forecast models.

Besides , Santhosh et al. [95] reported a comprehensive review of the recent advances and techniques for wind speed and wind power forecasting for improved integration of renewable energy. They covered a wide range of techniques including statistical models, artificial intelligence-based models (such as neural networks and support vector machines), and hybrid models. The authors concluded that the highest performance in terms of accuracy and reliability was achieved by using hybrid models that combined physical models and machine learning algorithms. They inferred that there is a need to enhance wind speed and wind power predictions to maximize the integration of renewable energies into power networks, thus minimizing operational expenses and enhancing grid stability.

Over and above that , Rajagopalan et al. [96]. conducted this study to predict wind power generation and analyze forecasting errors. They used Autoregressive Moving Average (ARMA) modeling in an effort to generate short-term wind power forecasts, As we can see in Figure 1 the Autoregressive Moving Average (ARMA) forecasting model . The result was that ARMA is able to detect trends and variations in wind power generation with relatively low forecasting errors. The authors concluded that the method can be used to improve the integration of wind power into power grids, and emphasized error analysis to make the forecast models optimal.

Beyond that Lima et al. [98] conducted a study with the objective of developing a meteorological–statistical model for short-term wind power prediction. They used techniques that combine meteorological information and statistical techniques to improve the accuracy of the predictions. The results obtained show that their model is capable of predicting the variability of wind power generation in short terms accurately, which is crucial for an optimal integration of wind energy into power grids. The authors conclude that the model possesses the ability to be a good tool for wind power producers and grid operators in reducing uncertainties of this power source's intermittent nature.

Over and above that C. Croonenbroeck et al. [99] conducted a comparative study of time series models for short- to medium-term wind power forecasting.

They employed a series of modeling approaches, from simple classical time series (e.g., ARIMA) to advanced techniques such as long-memory models (ARFIMA) and dynamic regression models. The results showed that ARFIMA and dynamic regression models possess greater forecast precision than the simpler models. The authors concluded that the choice of model depends strongly on particular wind data features and that the most appropriate models for wind power forecasting are those that are able to identify both short-scale variation and long-term trends.

This study recognizes the necessity of model adjustment to data specificities with the aim of improving forecast quality in wind energy.

Along those lines Gallego et al. [100] conducted a study of local wind direction and wind speed on the management of wind power fluctuation using precise reference towards very short-term forecast for off-shore wind farms. Authors employed statistical treatment and forecasting procedures to examine influences from local winds with varying capacity to produce winds. They found that wind direction and velocity have significant effects on affecting the variability in power output, especially in short time scales.

The findings reveal that inclusion of such local factors in the predictive models would improve the accuracy of wind power predictions significantly. The authors determine that local wind conditions must be taken into account in order to optimize the management and integration of wind power into grids, especially in offshore sites where weather conditions may be quite unstable.

In parallel Karaman et al .[101] contributed to the forecasting of wind power using machine learning models. Authors explored different machine learning methods for improving the performance of wind power forecasting using meteorological variables and historical values. It was concluded that machine learning methods, especially neural network-based and ensemble-based models, have superior performance over traditional methods. The authors concluded that the usage of these models allows energy resources to be managed better and the implementation of wind power in electricity grids to be increased. This article highlights the requirement for advanced machine learning algorithms for optimizing renewable energy production.

In a similar vein Azimi et al. [102]carried out this work with a goal of refining the accuracy in forecasting wind power. They introduced a hybrid approach using wavelet analysis techniques along with data mining instruments like artificial neural networks (ANNs) and support vector machines (SVMs). The approach aims at modeling both linear and nonlinear patterns in time series of wind power..

The results show that the proposed hybrid model outperforms traditional approaches in precision and reliability, wherein there is a significant reduction of prediction errors. The authors conclude that by coupling wavelet analysis and data mining techniques, they are able to model the complex variability of wind power better which is crucial in optimizing energy , system management on the basis of renewable energies.

Furthermore , Demolli et al. [103] conducted a study on wind power generation forecasting using daily wind speed data and machine learning models. They implemented different machine learning methods, including artificial neural networks (ANNs), support vector machines (SVMs), decision trees, and random forests, for wind power generation forecasting. The results showed that models based on ANN and SVM were most precise in predicting. The authors concluded that application of machine learning methods is highly effective for enhancing the accuracy of wind generation forecasting, and this is particularly critical for optimized integration of renewable energy and effective management of the power grid.

Chapter III. Research Methodology

This chapter outlines the research methodology used to predict wind power generation using various machine learning methods. The study is grounded on the data that is collected from a Senvion MM82 wind turbine located in France that has been recording wind power data at an interval of every 10 minutes since 2017. The approach emphasizes the importance of learning input and output variables as well as latent distribution and temporal relationships in the data.

To enhance the precision of prediction, several machine learning models are explored and contrasted, including Random Forest, Decision Tree, XGBoost, CatBoost, and Linear Regression. All these models are described in relation to their structure, advantages, and suitability for the regression task in wind power forecasting. With the utilization of these models, the most effective approach to scalable and precise forecasting can be ascertained based on the non-Gaussian and complicated nature of wind power data.

III.1 The input and the output parameters and their importance in predicting wind power

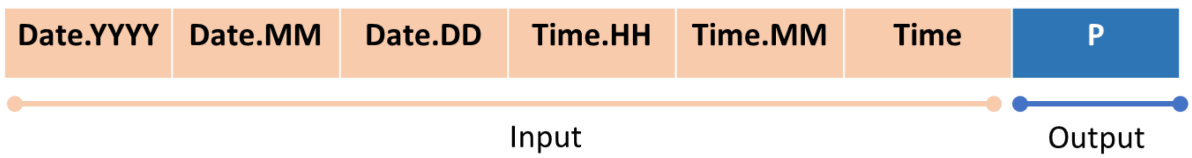
To determine the performance of the machine learning models proposed for wind power prediction, a dataset of measurements of wind power at a location in France has been utilized. A multi-site data collection approach increases the strength and applicability of the results in comparison to concentration on single sites or types of wind turbines. The French data was collected with a Senvion MM82 wind turbine that is 80 meters tall and has been generating data every 10 minutes since 2017.

The inputs are :

Date.YYYY (Year) , Date.MM(Month) , Date.DD (Day) , Time.HH (Hour) , Time.MM (Month) , Time (Total Time)

The output is:

P (Power) [Watts]



One of the most crucial steps in the analysis of wind power data is to investigate its underlying distribution and temporal dependencies therein. In ***Error! Reference source not found.***, violin p lots are shown for the distribution of wind power time-series data for a turbine located in France. These violin plots provide a complete picture by overlaying box plots, which indicate the median and interquartile range, with density plots, describing the overall spread and shape of the data. The finding from Figure 11 is an interesting observation: the wind power data is not of Gaussian distribution type; instead, it is of non-Gaussian nature. This indicates that the data deviates from the typical bell-shaped curve of normal distributions. Additionally, we note variations in both the distributions and shapes of data throughout France, indicating the potential influences of geographic features like prevailing winds and local topography in characterizing different wind regimes.[104]

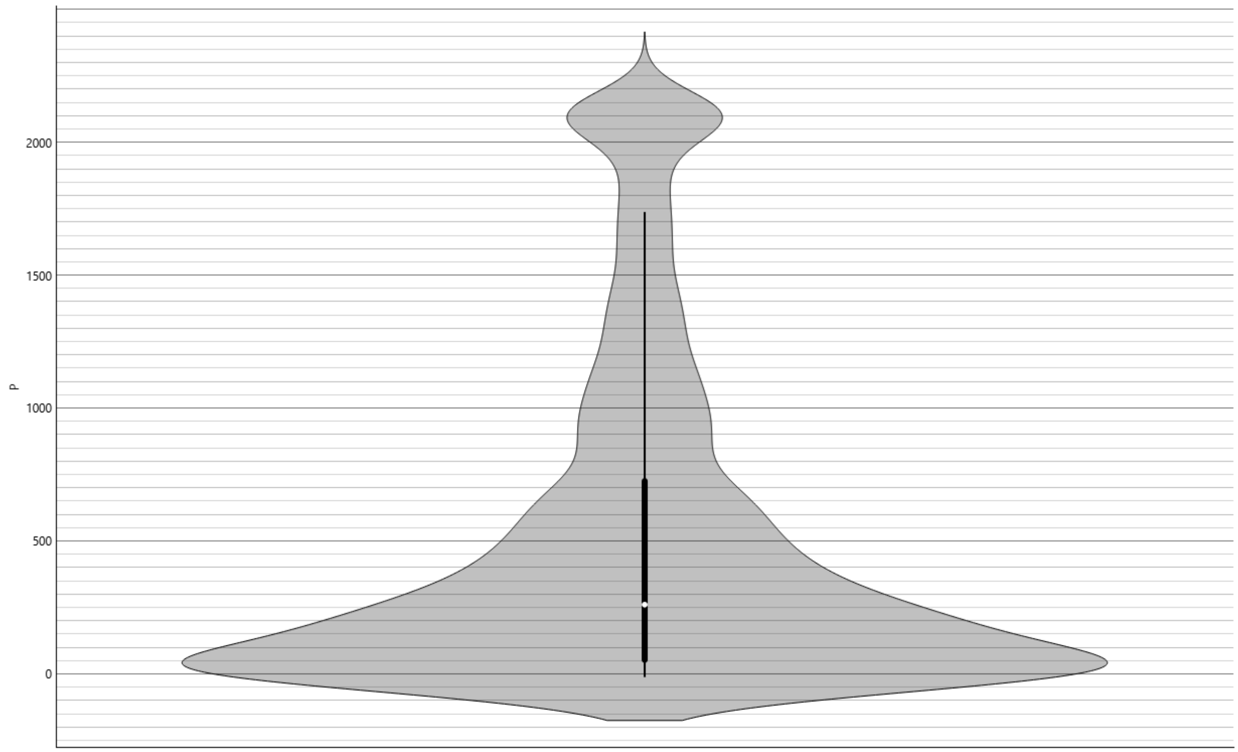


Figure 11 : Violin plots (a) of wind power time series data distribution for the wind turbine located in France [104]

III.2 The selection of Machine Learning models

III.2.1 Random forest

Random forest, or random decision forest, is an ensemble learning method which can be used for classification, regression, and other purposes. The algorithm builds many decision trees during training. For classification problems, the output of the random forest is the most voted class by the trees. For regression problems, the output is the average or mean prediction of the trees [105]. random decision forests address the issue of decision trees overfitting their training data.

Kam Ho [105] employs the random subspace method, which in his account is a method for implementing the "stochastic discrimination" approach to classification of Eugene Kleinberg.
Leo Breiman and Adele Cutler developed an extension of the algorithm in 2006 and patented "Random Forests." The algorithm extension consists of Breiman's concept of "bagging" with random feature selection, initially described by Ho and then separately by Amit and Geman. The outcome is an ensemble of variance-controlled decision trees.[106]

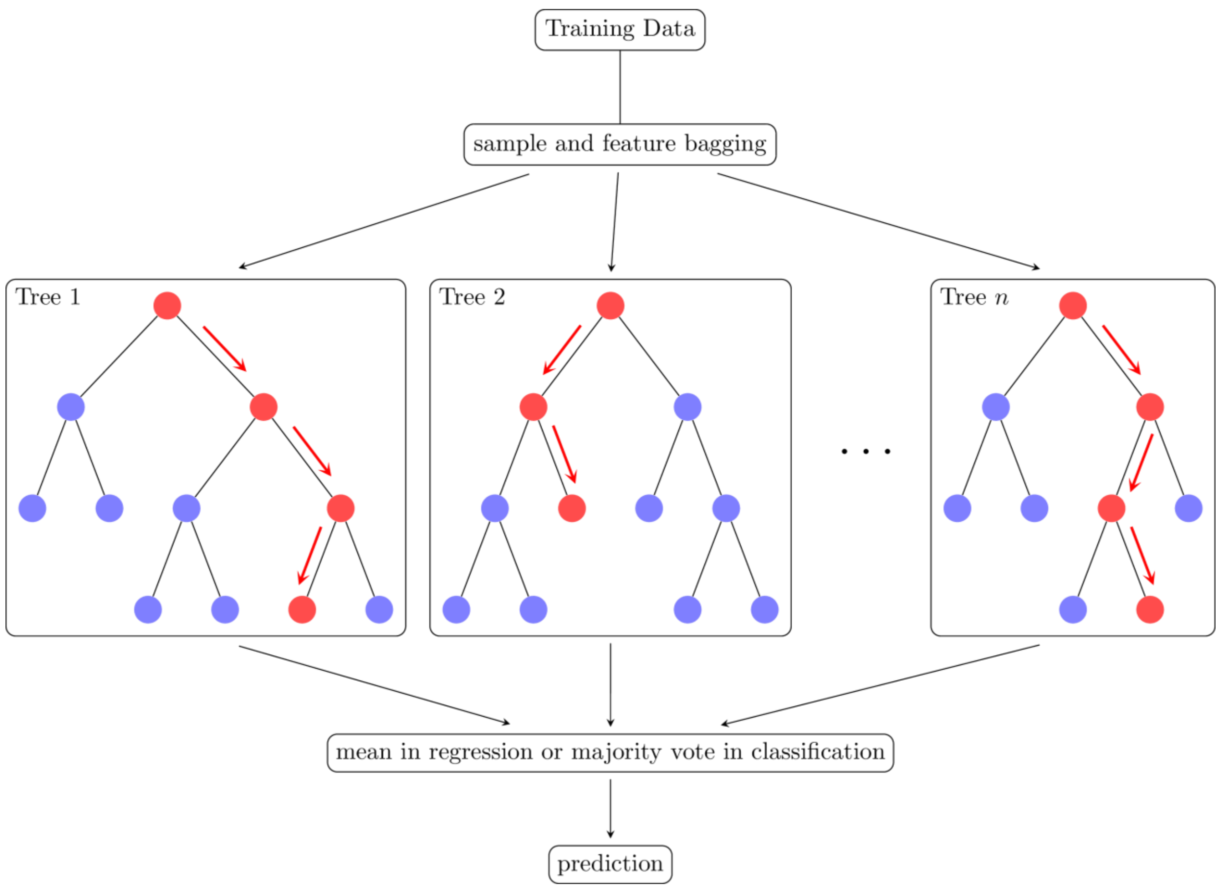


Figure 12 : Random forest [107].

III.2.2 Decision Tree

Decision tree learning is a supervised learning technique that is widely applied in statistics, data mining, and machine learning. Regression decision trees or decision trees are employed as models for prediction for classification or regression by this technique to make conclusions from an array of observations.

Classification trees are built in cases where the target variable can assume individual discrete values. In such tree models, class labels are indicated by leaves, and feature relations which lead to labels are given by the branches. Regression trees, in contrast, handle cases where the target variable can assume real values, typically referred to as real numbers. This concept may be further generalized to pairs of any type of objects that have pairwise dissimilarity, for instance, sequences of categories.

One of the most significant reasons why decision trees are a favorite among the machine learning community is that they are easy and straightforward to interpret [108]. Here's the diagram of decision tree in Figure 13

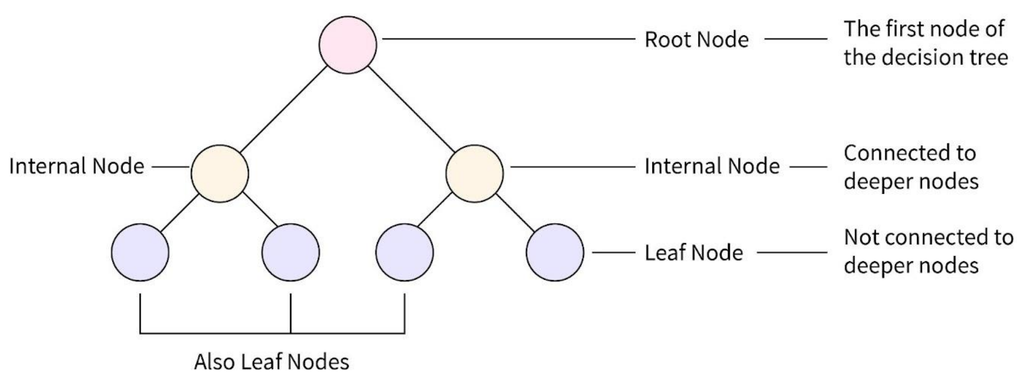


Figure 13: decision tree in machine learning

III.2.3 XGBoost

XGBoost or (eXtreme Gradient Boosting) is a distributed and open-source machine learning library. XGBoost makes use of gradient boosted decision trees, which is a supervised algorithm for learning with gradient descent being used to boost. XGBoost is said to be very efficient, fast, and also highly scalable for big data.

XGBoost is a highly optimized gradient boosting written by Tianqi Chen at the University of Washington. It is based on the same fundamental algorithm, which involves combining weak learner trees into strong learners through residual accumulation. The library has been implemented in many programming languages like C++, Python, R, Java, Scala, and Julia[109]

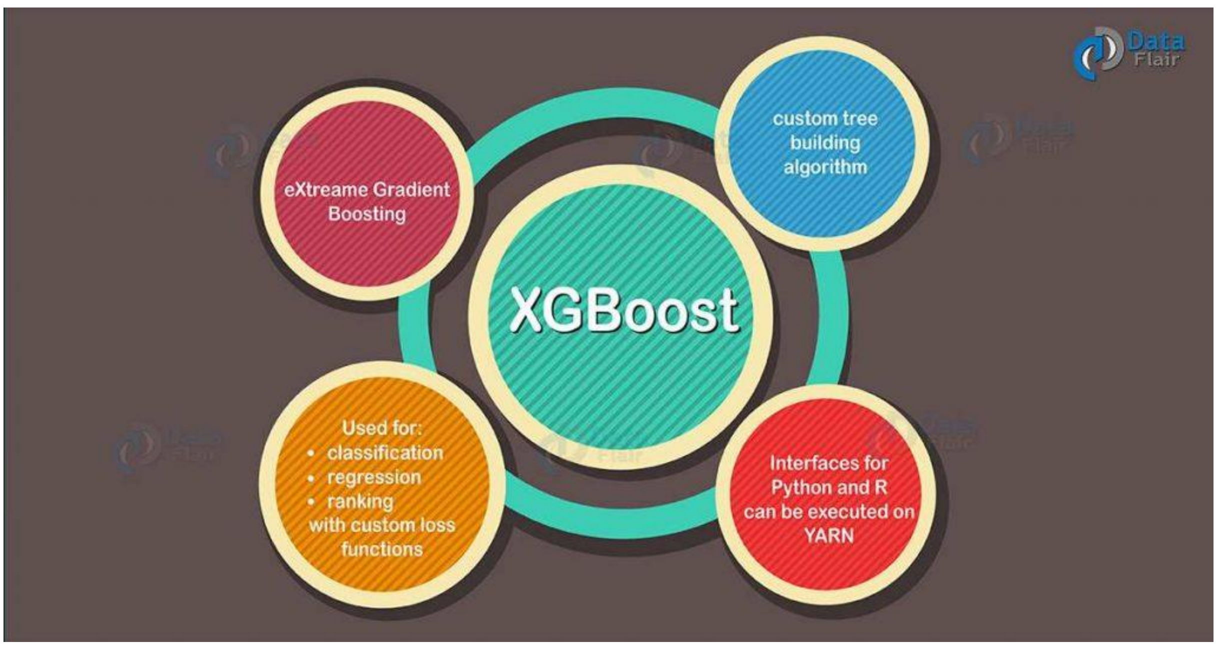


Figure 14 : XGBoost Algorithm – Objective.[110]

III.2.4 CatBoost

CatBoost is a gradient-boosting method of high performance specifically designed to deal with categorical features. CatBoost was constructed by Yandex and is highlighted for its ability to handle categorical data with minimal preprocessing required. It possesses exceptionally high accuracy during predictive modeling with minimal parameter adjustment required. It is optimized with implementation that facilitates it to train on large sets quickly. Due to its flexibility, overfitting resistance, and simplicity, CatBoost is a favorite among machine learning practitioners for regression and classification tasks. [111]

CatBoost utilizes a number of approaches that are employed to increase the accuracy and efficiency of the gradient boosting. They include decision tree optimization, feature engineering, and an optimized algorithm referred to as ordered boosting. [112]

With each algorithm iteration, CatBoost approximates the negative gradient of the loss function against the current prediction. The gradient is used for prediction updates by adding a scaled copy of the gradient to the existing predictions. The scaling factor is determined using a line search algorithm that aims to optimize the loss function. CatBoost builds decision trees using a process referred to as gradient-based optimization.[112]

In the process, trees are trained such that they would fit the negative gradient of the loss function. This allows trees to concentrate on the regions in the feature space that have the greatest impact on the loss function and make their predictions more efficiently. [112]

Here’s the flow diagram of the catboost model in Figure 15[113]

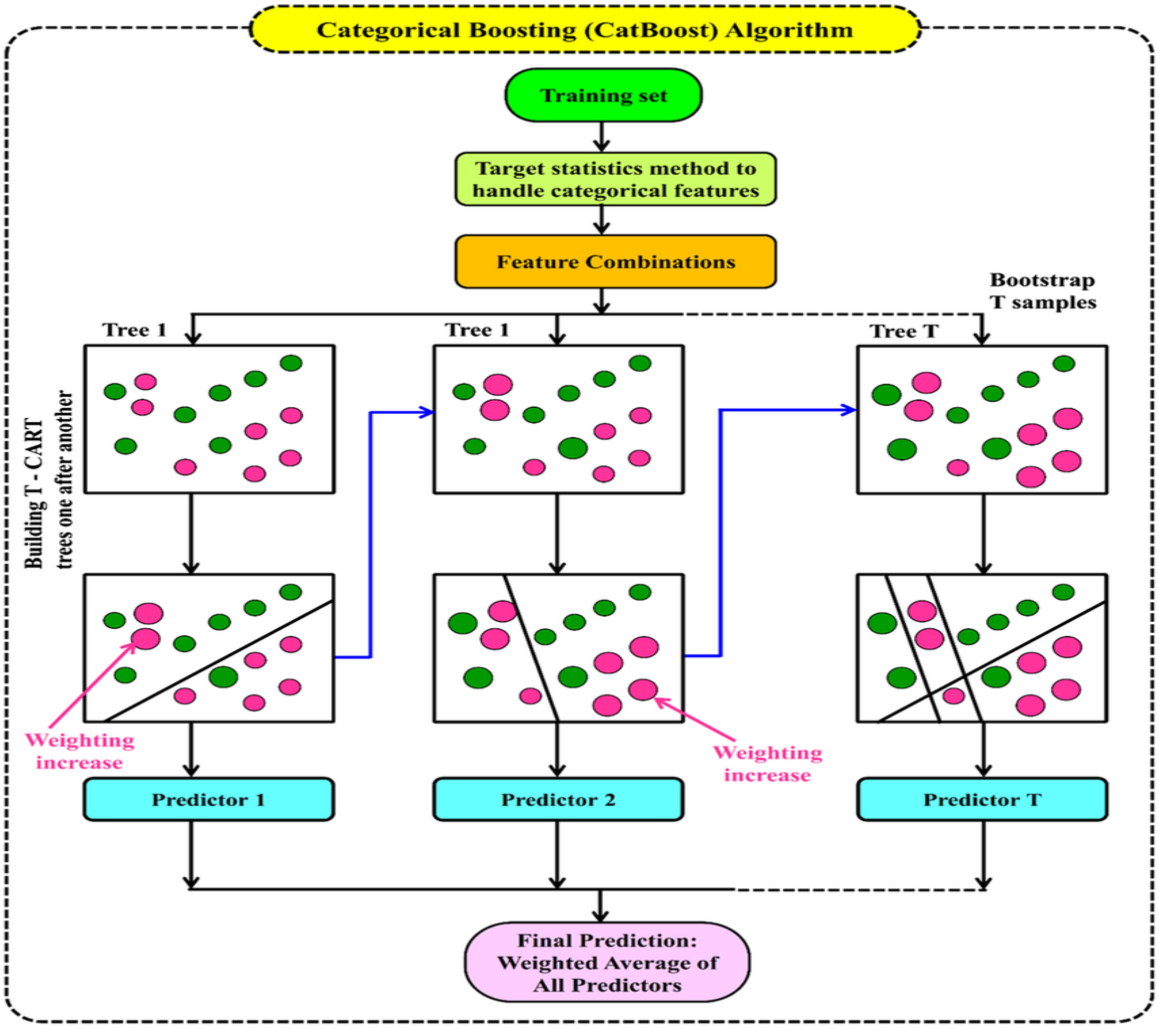


Figure 15: The flow diagram of the catboost model

III.2.5 Linear Regression

Linear regression is a fundamental technique widely used in statistical modeling and machine learning [114]. It is utilized to examine how an independent variable correlates with dependent variables, either single or multiple [115]. Linear regression is based on finding the line that best fits or runs through a collection of points, basically through minimizing the difference between forecast points and the actual points observed [116].

linear regression is a method models a linear equation, or a straight line, through the data set [117]. This line represents the best estimation of the dependent variable based on the values of the independent variables [115].

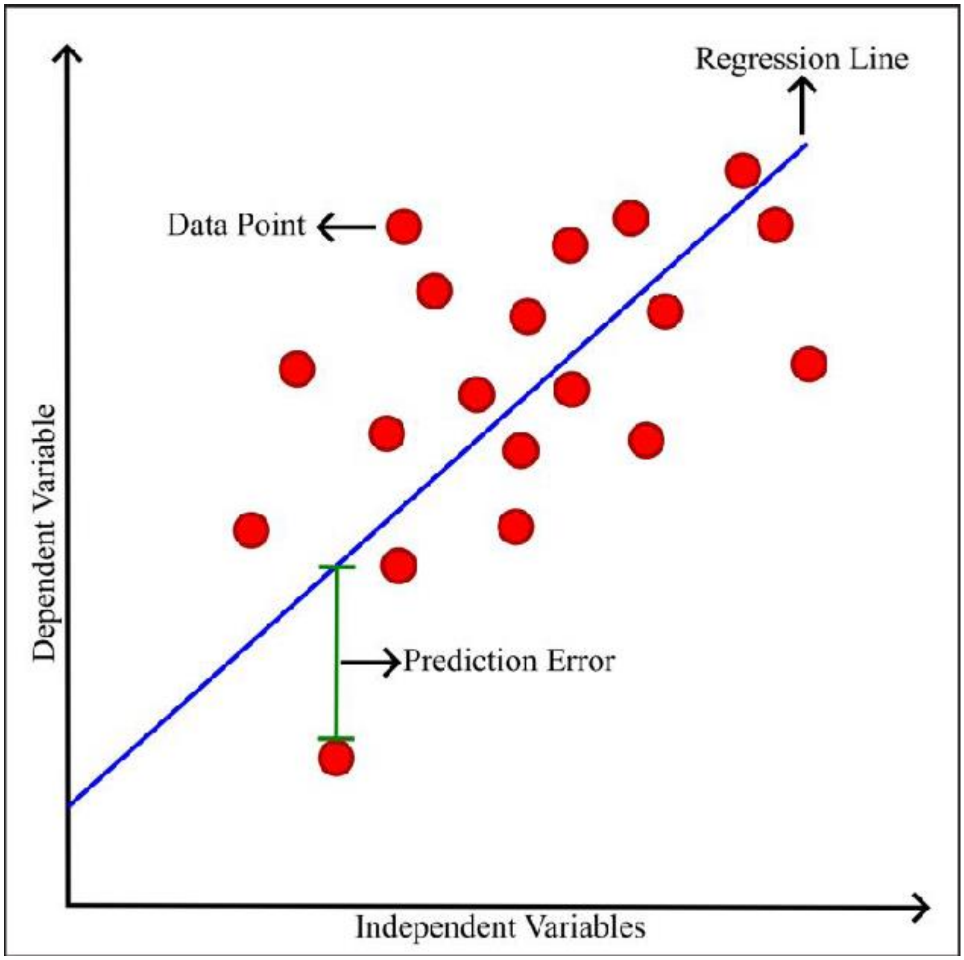


Figure 16: Linear Regression model [118]

Chapter IV. Simulation part

This chapter presents a thorough examination of machine learning models used for the prediction of wind power with focus on the measurement of the effect of various quantitative measures in judging model performance. It talks about significant evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2) and explains how they are used to measure prediction accuracy and model credibility. The chapter describes the approach of comparing different regression models like Linear Regression, Decision Tree, Random Forest, CatBoost, and XGBoost using a systematic experimental design approach. It also describes data preprocessing, model training, model testing, and how methods of lagged data are incorporated to enhance the predictive ability. The outcome and visualizations in terms of violin plots and comparison tables are deeply provided with the strengths and weaknesses of each of the models. In general, this chapter gives the foundation for selecting the optimal predictive model based on extensive performance assessment.

IV.1 Evaluation metrics (R^2, MAE , MAPE , RMSE,MSE)

In the final step, we continue with model evaluation by comparing predicted outcomes with real data. We use several metrics to assess performance, including the coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), and Mean Squared Error (MSE). The R^2 value indicates the goodness of fit of the model; the closer to 1, the more satisfactory the fit. On the other hand, we look for lower values of RMSE, MAE and MSE because these imply that the model makes more accurate predictions and has good performance.

- **R2** signifies the portion of variability in the target variable that can be anticipated based on the independent variables.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\boldsymbol{y}_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{y}_i)^2} \quad \text{(IV.1)}$$

- **RMSE** quantifies the disparity between predicted and actual values by considering the standard deviation of the error.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2} \quad \text{(IV.2)}$$

- **MAE** measures the absolute difference between the predicted and actual values.

$$MAE = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{n} \quad \text{(IV.3)}$$

- **MSE** stands for Mean Squared Error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \hat{y}_i)^2 \quad \text{(IV.4)}$$

where n represents the number of samples, y_t denotes the actual value of the response variable, and $\widehat{y_i}$ refers to the predicted value.

IV.2 Results and discussion

IV.2.1 Program presentation

This Figure 17 shows a screenshot of our machine learning workflow built in Orange, where our dataset is used to train and evaluate several regression models: XGBoost, CatBoost, Linear Regression, Decision Tree, and Random Forest. The models are assessed using the Test and Score widget to compare their performance. Additionally, a Data Table displays the dataset, and a Violin Plot visualizes feature distributions. This setup allows for efficient model comparison and data analysis.

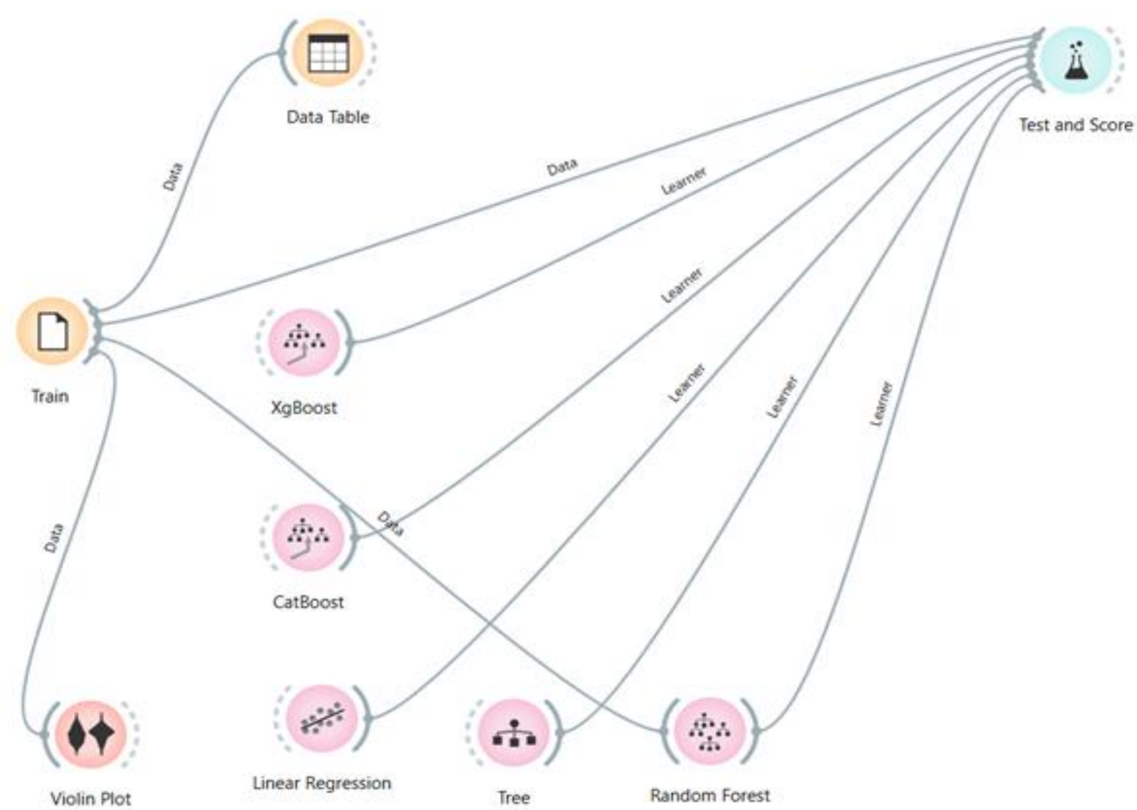


Figure 17: Screenshot of the simulation program of our study

We initially tried using orange for our study, but the outcomes were not satisfactory. Consequently, we switched to Python and applied a lagged data technique, which helped us achieve good results.

IV.2.2 Power (P) [Watts]

This violin plot presents a close-up of the power (P) value distribution of our study . The violin curve indicates the data point density at different levels of power. Most informative perhaps, the power distribution is skewed, with a very high density of power values at the lower end of the spectrum. This can be identified from the large base at the bottom of the plot, and this indicates that the majority of the observations have been made about very low levels of power.

As we move along the y-axis, the spread narrows considerably, and there are less data points toward the middle ranks of power. However, we do notice a bulge at the end of the plot, reflecting a second cluster of higher-value powers. The double form is typical of a bimodal distribution where the data clumps around two distinct ranges—one at low power and another at high power, with reduced numbers between them.

The white dot centered is the median power, and the thick black bar is the interquartile range (IQR), or the middle 50% of the data. The pale black line, drawn vertically, is the whole range from min to max values of power. The range highlights a few of the extreme values or outliers, which could potentially be at the very high levels, which may be instances of peak power utilization.

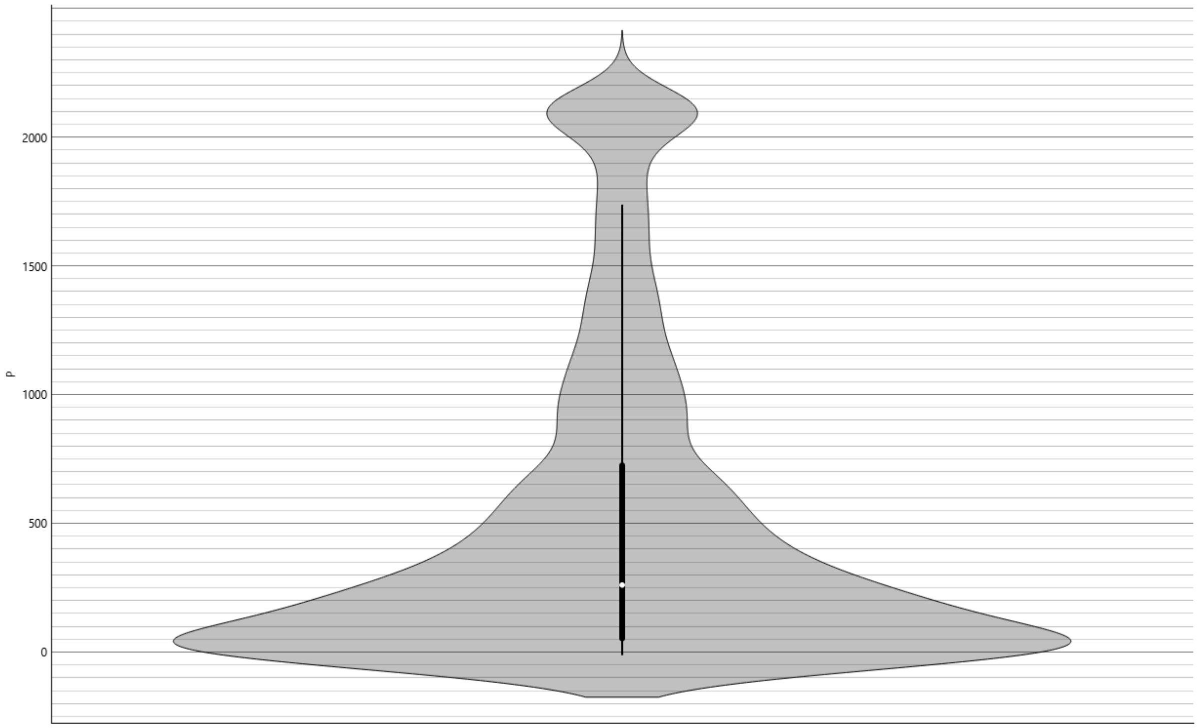


Figure 18 : Violin plot of Power(P)[Watts]

IV.3 Prediction framework

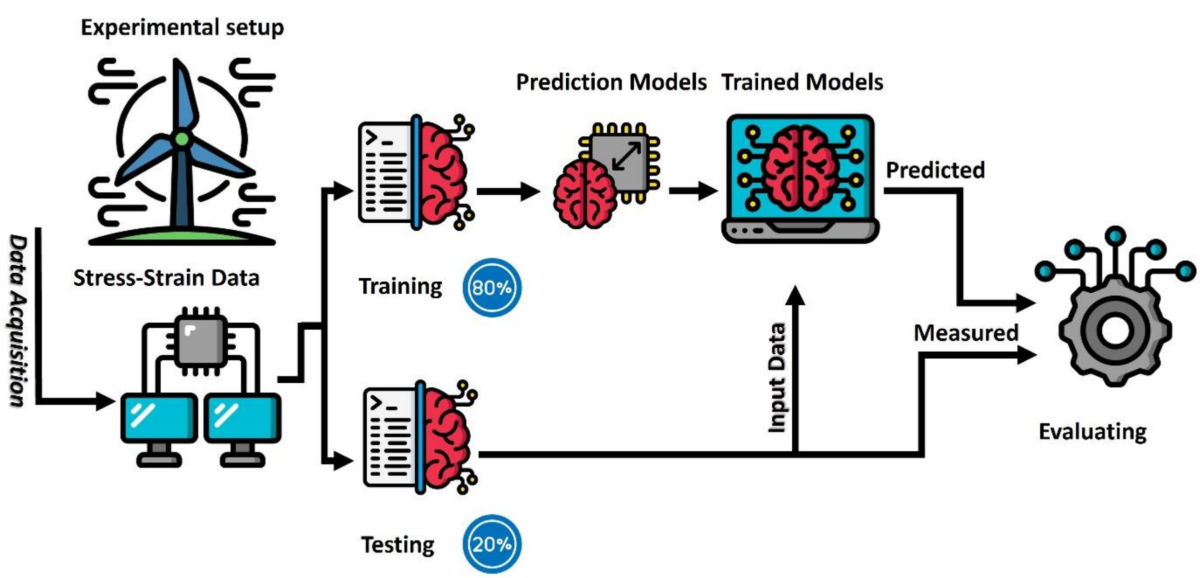


Figure 19 : Prediction framework of our study

Figure 19 presents the prediction framework of our study. It begins with experimental preparation, in which the theoretical or physical system—such as a wind turbine—is subjected to controlled conditions. This is to ensure that data collected is relevant, reliable, and consistent for purposes of prediction.

Following the installation, the data gathering process is collecting real-time or historical data from various sources, typically through sensors on the wind turbine or through environmental monitoring units. These sources may include wind speed, direction, temperature, and turbine power output, among others, and are all important for proper modeling.

Once data is collected, it is structured as input data. Preprocessing of the data happens in this stage, and such preprocessing may include cleaning (removing noise or outliers), normalization, and feature engineering to enhance the quality and usefulness of the dataset. These operations ensure that input is presented to the machine learning models such that efficient learning and generalization can be achieved.

The second phase involves training and testing. In this phase, machine learning models are trained on a portion of the dataset. The trained models are tested against another independent test set in an effort to gauge their performance. This division serves to prevent overfitting and guarantee that the models will perform well on new data.

Evaluating the prediction model is accomplished by matching predicted output against actual measured data. Various performance metrics, such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), or the Coefficient of Determination (R^2), are used to quantify the accuracy and robustness of the models. Finally, the results are tabulated in the predicted vs. measured comparison, providing visual or statistical information on how well the model performs, for further tuning or model selection.

Generally, this experimental arrangement ensures a straightforward, data-driven process from system configuration to accurate wind power forecasting, according to machine learning best practices.

IV.4 Lagged data technique

The Lagged Data Technique is a technique applied mainly in time series analysis and forecasting model building, where historical values of a variable are used as extra inputs (or features) in a model to assist in forecasting future or current values. The historical values, which are referred to as "lags," are indicative of how the variable performed at earlier time steps, e.g., one day ago, one week ago, or one month ago. Adding such lagged versions of a variable, analysts may detect temporal trends, patterns, or autocorrelation in the data, which typically are required for doing accurate predictions.

It is most common in models such as AR (AutoRegressive), ARIMA, and machine learning models operating with sequence data. For example, when forecasting today's temperature, one might use the temperature of the last one or few days as input features. Lagged data will have the ability to show patterns among past and current values to give information on how previous behavior will influence the future. Care should be taken to ensure that the data are ordered by time and addressed the missing values lagging often creates at the beginning of a dataset.

In the initial phase of our study, we utilized **Orange3**, a visual programming tool for data mining, to forecast wind power generation. However, the results obtained were suboptimal in terms of predictive accuracy and flexibility for advanced feature engineering. Consequently, we transitioned to a Python-based approach using **Jupyter Notebook**, which provided greater control over the modeling process. This allowed us to implement techniques such as **lag-based feature engineering**, which incorporates historical power output values to improve the performance of machine learning models.

To incorporate temporal dependencies into the model, we created a lag-based feature from the target variable P. Specifically, we used the previous time step's power value (P_lag1) as an input feature. This allows the models to learn from historical patterns. The lagging process was applied to both training and testing datasets, and the initial rows with missing values due to lagging were removed to ensure data integrity.

Below is the Lag Feature (Code Block) we used in our python code:

```
def add_lag_features(df, lags=[1]):
    for lag in lags:
        df[f'P_lag{lag}'] = df['P'].shift(lag)
    return df.dropna().reset_index(drop=True)

# Apply to both training and testing datasets
train_df = add_lag_features(train_df)
test_df = add_lag_features(test_df)
```

IV.5 Models evaluation

The Table 3 presents a comparison of the performance of five machine learning algorithms—Decision Tree, Random Forest, Linear Regression, CatBoost, and XGBoost—on four critical evaluation metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). The poorest of the five algorithms is the Decision Tree algorithm. It has the highest MSE (96227.19), i.e., it has the highest average squared difference between actual and predicted values. Similarly, its MAE (198.81) and RMSE (310.21) are also the greatest, indicating that the predictions of the model are quite far away from the actuals. Its R^2 is 0.5589, indicating that it only accounts for about 56% of variance in the data. Linear Regression works best overall. It has the smallest MSE (26209.56), RMSE (161.89), and MAE (106.75), which indicates that it is predicting more accurately. Its R^2 of 0.8799 also bears witness to this, the model explaining around 88% of the variance in the target variable. The Random Forest model is performing well too, the R^2 being 0.8519, a notch behind Linear Regression. Its measures of error (MSE = 32296.39, RMSE = 179.71, MAE = 120.81) are larger, which means that while it captures a high rate of variance, its individual predictions might be less precise than the Linear Regression model. CatBoost's performance is in between. It obtains an MSE of 49,757.47 and an R^2 of 0.7719, which shows greater predictive power than Decision Tree and XGBoost models but less than

Random Forest or Linear Regression. Its RMSE (223.06) and MAE (138.49) reflect medium-level predictive precision. XGBoost works slightly better than Decision Tree but still lags behind the rest of the models. It has an MSE of 74133.20, RMSE of 272.27, and MAE of 165.85. The R^2 value of 0.6602 shows that it explains about 66% of variance in the target variable. Finally, Linear Regression shows the best and most consistent performance of all the models according to all four of the measures, followed by Random Forest. CatBoost and XGBoost both show medium performance, and the lowest-performing model in the comparison is Decision Tree.

Table 3: Evaluation metrics of the used models

Model	MSE	RMSE	MAE	R^2
Decision Tree	96227.1871	310.2051	198.8136	0.5589
Random Forest	32296.3944	179.7120	120.8098	0.8519
Linear Regression	**26209.5602**	**161.8937**	**106.7462**	**0.8799**
CatBoost	49757.4659	223.0638	138.4926	0.7719
XgBoost	74133.2038	272.2741	165.8477	0.6602

The comparison of five machine learning algorithms reveals that **Linear Regression** achieved the best overall performance, with the lowest error values and the highest R^2 (0.8799), indicating strong predictive accuracy. **Random Forest** followed closely, explaining 85% of the variance, though with slightly higher error metrics. **CatBoost** and **XGBoost** delivered moderate performance, while the **Decision Tree** algorithm performed the worst, with the highest error rates and the lowest R^2 (0.5589), reflecting limited predictive capability. Overall, Linear Regression proved to be the most reliable model for forecasting wind power in this study.

Figure 20 presents a diagram of the evaluation metrics showing what was discussed previously.

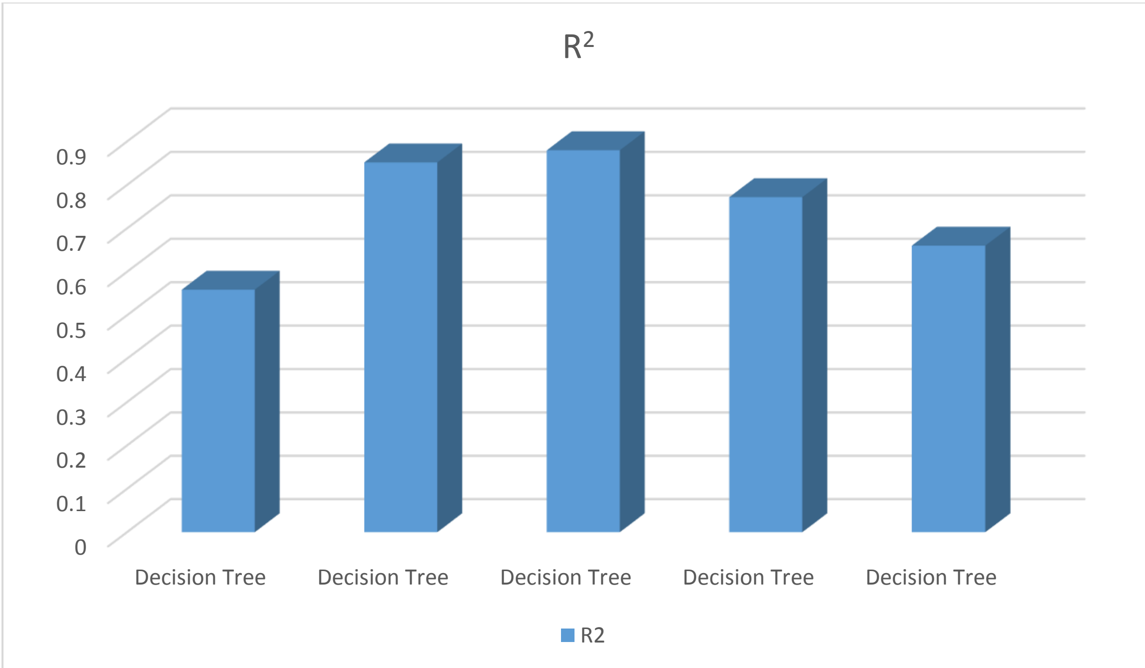


Figure 20 : Coefficient of determination evaluation metric of the machine learning models

General Conclusion

Our study discusses the application of machine learning algorithms to this complex task of predicting wind power, which is the core requirement for the holistic integration of renewable energy into today's power grids. Using real-time data collected from a Senvion MM82 wind turbine in France, five different models from the ML toolkit were trained and their performance evaluated with respect to several key performance metrics such as MAE, RMSE, MSE, and R^2.

Among all the models, Linear Regression performed optimally, delivering the best trade-off between simplicity and prediction accuracy. The article highlights the importance of proper data preprocessing, feature engineering, and inclusion of relevant meteorological and operational parameters to improve model performance. The lagged data methodology also played an important role in improving the accuracy of time-series prediction.

By comparative analysis, our study proves that machine learning, when used appropriately, presents a viable, scalable, and effective method for predicting wind power generation. This goes a long way towards ensuring the stability, reliability, and optimization of renewable energy systems.

The research is in line with the emerging consensus that hybrid models—fusing physical and data-driven methods—and real-time prediction systems can further increase wind power prediction and assist more intelligent energy management in the future.

Perspective

In the future, the study will be further pursued by exploring hybrid prediction models using several machine learning approaches to further improve accuracy in prediction. Furthermore, we aim to develop an application that can predict wind energy from real-time data, and this will be helpful tools for renewable energy operators and decision-makers.

References

[1] « History of wind power - U.S. Energy Information Administration (EIA) ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.eia.gov/energyexplained/wind/history-of-wind-power.php

[2] « Southeast Asia’s pathway to clean energy is all but clean | White paper », Hinrich Foundation. Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.hinrichfoundation.com/research/wp/sustainable/southeast-asia-pathway-to-clean-energy/

[3] H.-Y. Wang *et al.*, « Optimal wind energy generation considering climatic variables by Deep Belief network (DBN) model based on modified coot optimization algorithm (MCOA) », *Sustain. Energy Technol. Assess.*, vol. 53, p. 102744, oct. 2022, doi: 10.1016/j.seta.2022.102744.

[4] « WPA_2020Q3_Public_Version.pdf ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://cleanpower.org/wp-content/uploads/2021/02/WPA_2020Q3_Public_Version.pdf

[5] « General scheme of an onshore wind farm. | Download Scientific Diagram ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.researchgate.net/figure/General-scheme-of-an-onshore-wind-farm_fig3_368937951

[6] S. Hanifi, X. Liu, Z. Lin, et S. Lotfian, « A Critical Review of Wind Power Forecasting Methods—Past, Present and Future », *Energies*, vol. 13, nº 15, Art. nº 15, janv. 2020, doi: 10.3390/en13153764.

[7] « A review of wind power forecasting & prediction ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.researchgate.net/publication/311461324_A_review_of_wind_power_forecasting_prediction

[8] T. Ouyang, X. Zha, L. Qin, Y. He, et Z. Tang, « Prediction of wind power ramp events based on residual correction », *Renew. Energy*, vol. 136, nº C, p. 781-792, 2019.

[9] F. Ding, Z. Tian, F. Zhao, et H. Xu, « An integrated approach for wind turbine gearbox fatigue life prediction considering instantaneously varying load conditions », *Renew. Energy*, vol. 129, nº PA, p. 260-270, 2018.

[10] « (PDF) Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.researchgate.net/publication/330998114_Mid-to-long_term_wind_and_photovoltaic_power_generation_prediction_based_on_copula_function_and_long_short_term_memory_network

[11] « A review of combined approaches for prediction of short-term wind speed and power | Request PDF », *ResearchGate*, oct. 2024, doi: 10.1016/j.rser.2014.03.033.

[12] B. Bouyeddou, F. Harrou, A. Saidi, et Y. Sun, « An Effective Wind Power Prediction using Latent Regression Models », in *2021 International Conference on ICT for Smart Society (ICISS)*, Bandung, Indonesia: IEEE, août 2021, p. 1-6. doi: 10.1109/ICISS53185.2021.9533242.

[13] M. Khezri, M. S. Karimi, J. Mamkhezri, R. Ghazal, et L. Blank, « Assessing the Impact of Selected Determinants on Renewable Energy Sources in the Electricity Mix: The Case of ASEAN Countries », *Energies*, vol. 15, nº 13, Art. nº 13, janv. 2022, doi: 10.3390/en15134604.

[14] C. C. Sousa, J. Martins, Ó. Carvalho, M. Coelho, A. S. Moita, et F. P. Brito, « Assessment of an Exhaust Thermoelectric Generator Incorporating Thermal Control Applied to a Heavy Duty Vehicle », *Energies*, vol. 15, n° 13, Art. n° 13, janv. 2022, doi: 10.3390/en15134787.

[15] « Prospects of Fuel Cell Combined Heat and Power Systems ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/13/16/4104

[16] « Exploring the Causal Relationship among Green Taxes, Energy Intensity, and Energy Consumption in Nordic Countries: Dumitrescu and Hurlin Causality Approach ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/15/14/5199

[17] « Optimising the Concentrating Solar Power Potential in South Africa through an Improved GIS Analysis ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/13/12/3258

[18] « Study on the Motion Characteristics of 10 MW Superconducting Floating Offshore Wind Turbine Considering 2nd Order Wave Effect ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/14/19/6070

[19] « Energy and Nutrients' Recovery in Anaerobic Digestion of Agricultural Biomass: An Italian Perspective for Future Applications ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/12/17/3287

[20] K.-W. Ng, W.-H. Lam, et K.-C. Ng, « 2002–2012: 10 Years of Research Progress in Horizontal-Axis Marine Current Turbines », *Energies*, vol. 6, n° 3, Art. n° 3, mars 2013, doi: 10.3390/en6031497.

[21] B. R. Kumar, « Case 16: Three Gorges Dam—The World's Largest Hydroelectric Plant », *Manag. Prof.*, p. 183-186, 2022.

[22] « ITAIPU BINACIONAL | World leader in clean and renewable energy ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.itaipu.gov.br/en

[23] B. R. Kumar, « Case 21: Bhadla Solar Park », Springer, Management for Professionals, 2022. Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://econpapers.repec.org/bookchap/sprmgmchp/978-3-030-96725-3_5f25.htm

[24] Z. Yang *et al.*, « Sizing utility-scale photovoltaic power generation for integration into a hydropower plant considering the effects of climate change: A case study in the Longyangxia of China », *Energy*, vol. 236, n° C, 2021, Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://ideas.repec.org//a/eee/energy/v236y2021ics0360544221017679.html

[25] K. Beladjine, F. Sadji, et R. Beladjine, « China Experience in Renewable Energies », *مجلة افاق علوم الادارة والاقتصاد*, vol. 6, n° 1, p. 700-718, juin 2022.

[26] « Optimal wind energy generation considering climatic variables by Deep Belief network (DBN) model based on modified coot optimization algorithm (MCOA) ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.researchgate.net/publication/363505932_Optimal_wind_energy_generation_considering_climatic_variables_by_Deep_Belief_network_DBN_model_based_on_modified_coot_optimization_algorithm_MCOA

[27] T. Haidi et B. Cheddadi, « STATE OF WIND ENERGY IN THE WORLD: EVOLUTION, IMPACTS AND PERSPECTIVES », vol. 14, n° 2, 2022.

[28] « (PDF) Decision Making Of Suitable Bioenergy Power Plant Location: A Case Study », in *ResearchGate*, doi: 10.15439/2020KM32.

[29] Yumpu.com, « The worldÂ´s largest biofuel CHP plant Alholmens Kraft, Pietarsaari(pdf) », yumpu.com. Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.yumpu.com/en/document/view/27538435/the-worldas-largest-biofuel-chp-plant-alholmens-kraft-pietarsaaripdf

[30] T. Simla et W. Stanek, « Influence of the wind energy sector on thermal power plants in the Polish energy system », *Renew. Energy*, vol. 161, nº C, p. 928-938, 2020.

[31] S. A. Awuku, A. Bennadji, F. Muhammad-Sukki, et N. Sellami, « Promoting the Solar Industry in Ghana through Effective Public-Private Partnership (PPP): Some Lessons from South Africa and Morocco », *Energies*, vol. 15, nº 1, Art. nº 1, janv. 2022, doi: 10.3390/en15010017.

[32] « Ivanpah Solar Electric Generating System | Concentrating Solar Power Projects | NREL ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://solarpaces.nrel.gov/project/ivanpah-solar-electric-generating-system

[33] « Mojave Solar Project | Concentrating Solar Power Projects | NREL ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://solarpaces.nrel.gov/project/mojave-solar-project

[34] « Heating Sizing Power Reduction in Buildings Connected to District Heating with Dynamically Controlled DHW Setback and Flow Limiters ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/15/14/5278

[35] « Power Converter Topologies for Grid-Tied Solar Photovoltaic (PV) Powered Electric Vehicles (EVs)—A Comprehensive Review ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/15/13/4648

[36] « State-of-the-Art Technologies for Building-Integrated Photovoltaic Systems ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/2075-5309/11/9/383

[37] « Bibliometric Analysis of Solar Desalination Systems Powered by Solar Energy and CFD Modelled ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/15/14/5279

[38] Y. B. Muna et C.-C. Kuo, « Feasibility and Techno-Economic Analysis of Electric Vehicle Charging of PV/Wind/Diesel/Battery Hybrid Energy System with Different Battery Technology », *Energies*, vol. 15, nº 12, Art. nº 12, janv. 2022, doi: 10.3390/en15124364.

[39] « Analysis of the Heat Transfer Coefficient, Thermal Effusivity and Mathematical Modelling of Drying Kinetics of a Partitioned Single Pass Low-Cost Solar Drying of Cocoyam Chips with Economic Assessments ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/15/12/4457

[40] « Determinants and Impact of Solar Irrigation Facility (SIF) Adoption: A Case Study in Northern Bangladesh ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/15/7/2460

[41] H. L. Zhang, J. Baeyens, J. Degrève, et G. Cacères, « Concentrated solar power plants: Review and design methodology », *Renew. Sustain. Energy Rev.*, vol. 22, p. 466-481, juin 2013, doi: 10.1016/j.rser.2013.01.032.

[42] « Technology Roadmap - Concentrating Solar Power – Analysis », IEA. Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.iea.org/reports/technology-roadmap-concentrating-solar-power

[43] M. Xu et C. Sheng, « Modeling the Process and Properties of Ash Formation during Pulverized Biomass Combustion », *Energies*, vol. 15, nº 12, Art. nº 12, janv. 2022, doi: 10.3390/en15124417.

[44] T. N. T. Pham et Y. S. Yoon, « Development of Nanosized Mn3O4-Co3O4 on Multiwalled Carbon Nanotubes for Cathode Catalyst in Urea Fuel Cell », *Energies*, vol. 13, nº 9, Art. nº 9, janv. 2020, doi: 10.3390/en13092322.

[45] N. Yang *et al.*, « Combustion Performance and Emission Characteristics of Marine Engine Burning with Different Biodiesel », *Energies*, vol. 15, nº 14, Art. nº 14, janv. 2022, doi: 10.3390/en15145177.

[46] « Pyrolysis of Biomass Wastes into Carbon Materials ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/15/5/1941

[47] « Ocean Wave Energy: Current Status and Future Prespectives | SpringerLink ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://link.springer.com/book/10.1007/978-3-540-74895-3

[48] « Rights and ownership in sea country: implications of marine renewable energy for indigenous and local communities ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://ideas.repec.org/a/eee/marpol/v52y2015icp108-115.html

[49] T. Ackermann et L. Söder, « An overview of wind energy-status 2002 », *Renew. Sustain. Energy Rev.*, vol. 6, n° 1, p. 67-127, janv. 2002, doi: 10.1016/S1364-0321(02)00008-4.

[50] « Wind energy: Trends and enabling technologies ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://ideas.repec.org/a/eee/rensus/v53y2016icp209-224.html

[51] « (PDF) Comparative Study on Horizontal Axis Wind Turbines and Vertical Axis Wind Turbines », ResearchGate. Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.researchgate.net/publication/351233790_Comparative_Study_on_Horizontal_Axis_Wind_Turbines_and_Vertical_Axis_Wind_Turbines

[52] « Selection Guidelines for Wind Energy Technologies ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/14/11/3244

[53] L. H. Hansen, P. H. Madsen, F. Blaabjerg, H. C. Christensen, U. Lindhard, et K. Eskildsen, « Generators and power electronics technology for wind turbines: 27th Annual Conference of the IEEE Industrial Electronics Society », *Proc. CD-ROM*, p. 2000-2005, 2001.

[54] T. Haidi et B. Cheddadi, « STATE OF WIND ENERGY IN THE WORLD: EVOLUTION, IMPACTS AND PERSPECTIVES », vol. 14, n° 2, 2022.

[55] « Practices and Challenge on Planning with Large-scale Renewable Energy Grid Integration », in *ResearchGate*, déc. 2024. doi: 10.1109/EI247390.2019.9062091.

[56] E. T. Sayed *et al.*, « Renewable Energy and Energy Storage Systems », *Energies*, vol. 16, n° 3, Art. n° 3, janv. 2023, doi: 10.3390/en16031415.

[57] « GWEC in 2020 », Global Wind Energy Council. Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://gwec.net/gwec-in-2020/

[58] « Renewable Energy Sources and Climate Change Mitigation — IPCC ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.ipcc.ch/report/renewable-energy-sources-and-climate-change-mitigation/

[59] « Future of wind ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.irena.org/publications/2019/Oct/Future-of-wind

[60] « Wind Energy Development and Wildlife Conservation: Challenges and Opportunities - KUVLESKY - 2007 - The Journal of Wildlife Management - Wiley Online Library ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://wildlife.onlinelibrary.wiley.com/doi/abs/10.2193/2007-248

[61] « Wind Energy Explained | Wiley Online Books ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119994367

[62] « The intermittency of wind, solar, and renewable electricity generators: Technical barrier or rhetorical excuse? | Request PDF ». Consulté le: 8 février 2025. [En ligne]. Disponible sur: https://www.researchgate.net/publication/46495038_The_intermittency_of_wind_solar_and_renewable_electricity_generators_Technical_barrier_or_rhetorical_excuse

[63] R. Wiser *et al.*, « 2018 Wind Technologies Market Report », 2018.

[64] A. Niayifar et F. Porté-Agel, « A new analytical model for wind farm power prediction », *J. Phys. Conf. Ser.*, vol. 625, n° 1, p. 012039, juin 2015, doi: 10.1088/1742-6596/625/1/012039.

[65] « Wind Energy Prediction Using Machine Learning ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.scirp.org/journal/paperinformation?paperid=106810

[66] A.-N. Buturache et S. Stancu, « Wind Energy Prediction Using Machine Learning », *Low Carbon Econ.*, vol. 12, nº 1, Art. nº 1, janv. 2021, doi: 10.4236/lce.2021.121001.

[67] « Comparison Between Wind Power Prediction Models Based on Wavelet Decomposition with Least-Squares Support Vector Machine (LS-SVM) and Artificial Neural Network (ANN) ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/7/8/5251

[68] A.-N. Buturache et S. Stancu, « Wind Energy Prediction Using Machine Learning », *Low Carbon Econ.*, vol. 12, nº 1, Art. nº 1, janv. 2021, doi: 10.4236/lce.2021.121001.

[69] Y. Liu, Y. Sun, D. Infield, Y. Zhao, S. Han, et J. Yan, « A Hybrid Forecasting Method for Wind Power Ramp Based on Orthogonal Test and Support Vector Machine (OT-SVM) », *IEEE Trans. Sustain. Energy*, vol. 8, nº 2, p. 451-457, avr. 2017, doi: 10.1109/TSTE.2016.2604852.

[70] A. Torres-Barrán, Á. Alonso, et J. R. Dorronsoro, « Regression tree ensembles for wind energy and solar radiation prediction », *Neurocomputing*, vol. 326-327, p. 151-160, janv. 2019, doi: 10.1016/j.neucom.2017.05.104.

[71] A. Lahouar et J. Ben Hadj Slama, « Hour-ahead wind power forecast based on random forests », *Renew. Energy*, vol. 109, p. 529-541, août 2017, doi: 10.1016/j.renene.2017.03.064.

[72] « Using Random Forests to Select Optimal Input Variables for Short-Term Wind Speed Forecasting Models ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/10/10/1522

[73] « Snapshot ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://builtin.com/data-science/curse-dimensionality

[74] A.-N. Buturache et S. Stancu, « Wind Energy Prediction Using Machine Learning », *Low Carbon Econ.*, vol. 12, nº 1, Art. nº 1, janv. 2021, doi: 10.4236/lce.2021.121001.

[75] « IEEE Xplore Abstract Record ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://ieeexplore.ieee.org/document/932320

[76] « IEEE Xplore Abstract Record ». Consulté le: 9 février 2025. [En ligne]. Disponible sur: https://ieeexplore.ieee.org/document/5749029

[77] J. P. S. Catalão, H. M. I. Pousinho, et V. M. F. Mendes, « Short-term wind power forecasting in Portugal by neural networks and wavelet transform », *Renew. Energy*, vol. 36, nº 4, p. 1245-1251, avr. 2011, doi: 10.1016/j.renene.2010.09.016.

[78] P. Flores, A. Tapia, et G. Tapia, « Application of a control algorithm for wind speed prediction and active power generation », *Renew. Energy*, vol. 30, nº 4, p. 523-536, avr. 2005, doi: 10.1016/j.renene.2004.07.015.

[79] V. Singh, « Application of Artificial Neural Networks for Predicting Generated Wind Power », *Int. J. Adv. Comput. Sci. Appl. Ijacsa*, vol. 7, nº 3, Art. nº 3, 31 2016, doi: 10.14569/IJACSA.2016.070336.

[80] « Forecasting Wind Patterns Using Machine Learning | Capitol Technology University ». Consulté le: 21 juin 2025. [En ligne]. Disponible sur: https://www.captechu.edu/blog/forecasting-wind-patterns-using-machine-learning

[81] O. Karakus, E. E. Kuruoglu, et M. A. Altinkaya, « One-day ahead wind speed/power prediction based on polynomial autoregressive model », *IET Renew. Power Gener.*, vol. 11, nº 11, p. 1430-1439, sept. 2017.

[82] S. Rajagopalan et S. Santoso, « Wind power forecasting and error analysis using the autoregressive moving average modeling », in *2009 IEEE Power & Energy Society General Meeting*, juill. 2009, p. 1-6. doi: 10.1109/PES.2009.5276019.

[83] « Wind Power Forecasting Using Hybrid ARIMA-ANN Technique | Request PDF », *ResearchGate*, nov. 2024, doi: 10.1007/978-981-13-5934-7_19.

[84] « Fig. 2.8. Block diagram of ARIMA(0,1,1) model of wind power time series. », ResearchGate. Consulté le: 14 février 2025. [En ligne]. Disponible sur:

https://www.researchgate.net/figure/Block-diagram-of-ARIMA0-1-1-model-of-wind-power-time-series_fig2_267827916

[85] J. Nielson, K. Bhaganagar, R. Meka, et A. Alaeddini, « Using atmospheric inputs for Artificial Neural Networks to improve wind turbine power prediction », *Energy*, vol. 190, p. 116273, janv. 2020, doi: 10.1016/j.energy.2019.116273.

[86] A. Alkesaiberi, F. Harrou, et Y. Sun, « Efficient Wind Power Prediction Using Machine Learning Methods: A Comparative Study », *Energies*, vol. 15, nº 7, Art. nº 7, janv. 2022, doi: 10.3390/en15072327.

[87] U. Singh et M. Rizwan, « SCADA system dataset exploration and machine learning based forecast for wind turbines », *Results Eng.*, vol. 16, p. 100640, déc. 2022, doi: 10.1016/j.rineng.2022.100640.

[88] « Wind Power Prediction Using Ensemble Learning-Based Models | IEEE Journals & Magazine | IEEE Xplore ». Consulté le: 16 février 2025. [En ligne]. Disponible sur: https://ieeexplore.ieee.org/document/9046858

[89] C. Cakiroglu, S. Demir, M. H. Ozdemir, B. L. Aylak, G. Sariisik, et L. Abualigah, « Data-driven interpretable ensemble learning methods for the prediction of wind turbine power incorporating SHAP analysis », *Expert Syst. Appl.*, vol. 237, nº A, mars 2024, doi: 10.1016/j.eswa.2023.121464.

[90] « Short-Term Power Prediction of Wind Turbine Applying Machine Learning and Digital Filter ». Consulté le: 16 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/2076-3417/13/3/1751

[91] H. Jiajun, Y. Chuanjin, L. Yongle, et X. Huoyue, « Ultra-short term wind prediction with wavelet transform, deep belief network and ensemble learning », *Energy Convers. Manag.*, vol. 205, p. 112418, févr. 2020, doi: 10.1016/j.enconman.2019.112418.

[92] C. Wang, H. Zhang, et P. Ma, « Wind power forecasting based on singular spectrum analysis and a new hybrid Laguerre neural network », *Appl. Energy*, vol. 259, p. 114139, févr. 2020, doi: 10.1016/j.apenergy.2019.114139.

[93] « F. Harrou, Y. Sun, A. Dairi, B. Taghezouit, S. Khadraoui, Advanced data-driven meth- ods for monitoring solar and wind energy systems, Front. Energy Res. 11 (2023) 1147746. - Recherche Google ». Consulté le: 15 février 2025. [En ligne]. Disponible sur: https://www.google.com/search?q=F.+Harrou%2C+Y.+Sun%2C+A.+Dairi%2C+B.+Taghezouit%2C+S.+Khadraoui%2C+Advanced+data-driven+meth-+ods+for+monitoring+solar+and+wind+energy+systems%2C+Front.+Energy+Res.+11+(2023)+1147746.&rlz=1C1CHNY_frDZ999DZ999&oq=F.+Harrou%2C+Y.+Sun%2C+A.+Dairi%2C+B.+Taghezouit%2C+S.+Khadraoui%2C+Advanced+data-driven+meth-+ods+for+monitoring+solar+and+wind+energy+systems%2C+Front.+Energy+Res.+11+(2023)+1147746.&aqs=chrome.0.69i59.1527j0j7&sourceid=chrome&ie=UTF-8

[94] « A Critical Review of Wind Power Forecasting Methods—Past, Present and Future ». Consulté le: 15 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/1996-1073/13/15/3764

[95] « (PDF) Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: A review ». Consulté le: 15 février 2025. [En ligne]. Disponible sur: https://www.researchgate.net/publication/341316610_Current_advances_and_approaches_in_wind_speed_and_wind_power_forecasting_for_improved_renewable_energy_integration_A_review

[96] S. Rajagopalan et S. Santoso, « Wind power forecasting and error analysis using the autoregressive moving average modeling », in *2009 IEEE Power & Energy Society General Meeting*, juill. 2009, p. 1-6. doi: 10.1109/PES.2009.5276019.

[97] « Fig. 3. Autoregressive moving-average (ARMA) predictive model:... », ResearchGate. Consulté le: 15 février 2025. [En ligne]. Disponible sur: https://www.researchgate.net/figure/Autoregressive-moving-average-ARMA-predictive-model-ARMA1-0-model_fig1_306285926

[98] « (PDF) A Meteorological–Statistic Model for Short-Term Wind Power Forecasting », *ResearchGate*, déc. 2024, doi: 10.1007/s40313-017-0329-8.

[99] C. Croonenbroeck et D. Ambach, « A selection of time series models for short- to medium-term wind power forecasting », *J. Wind Eng. Ind. Aerodyn.*, vol. 136, p. 201-210, janv. 2015, doi: 10.1016/j.jweia.2014.11.014.

[100] C. Gallego, P. Pinson, H. Madsen, A. Costa, et A. Cuerva, « Influence of local wind speed and direction on wind power dynamics – Application to offshore very short-term forecasting », *Appl. Energy*, vol. 88, nº 11, p. 4087-4096, nov. 2011, doi: 10.1016/j.apenergy.2011.04.051.

[101] « Prediction of Wind Power with Machine Learning Models ». Consulté le: 15 février 2025. [En ligne]. Disponible sur: https://www.mdpi.com/2076-3417/13/20/11455

[102] « A hybrid wind power forecasting model based on data mining and wavelets analysis | Request PDF », *ResearchGate*, oct. 2024, doi: 10.1016/j.enconman.2016.09.002.

[103] « Wind power forecasting based on daily wind speed data using machine learning algorithms | Request PDF », *ResearchGate*, déc. 2024, doi: 10.1016/j.enconman.2019.111823.

[104] F. Harrou, A. Dairi, A. Dorbane, et Y. Sun, « Enhancing wind power prediction with self-attentive variational autoencoders: A comparative study », *Results Eng.*, vol. 23, p. 102504, sept. 2024, doi: 10.1016/j.rineng.2024.102504.

[105] « Wayback Machine ». Consulté le: 17 mars 2025. [En ligne]. Disponible sur: https://web.archive.org/

[106] L. Breiman, « Random Forests », *Mach. Learn.*, vol. 45, nº 1, p. 5-32, oct. 2001, doi: 10.1023/A:1010933404324.

[107] « Structure of Random Forest [11]. | Download Scientific Diagram ». Consulté le: 17 mars 2025. [En ligne]. Disponible sur: https://www.researchgate.net/figure/Structure-of-Random-Forest-11_fig1_371946877

[108] X. Wu *et al.*, « Top 10 algorithms in data mining », *Knowl. Inf. Syst.*, vol. 14, nº 1, p. 1-37, janv. 2008, doi: 10.1007/s10115-007-0114-2.

[109] « What is XGBoost? | IBM ». Consulté le: 5 avril 2025. [En ligne]. Disponible sur: https://www.ibm.com/think/topics/xgboost

[110] D. Team, « What is XGBoost Algorithm - Applied Machine Learning », DataFlair. Consulté le: 5 avril 2025. [En ligne]. Disponible sur: https://data-flair.training/blogs/xgboost-algorithm/

[111] « CatBoost in Machine Learning - GeeksforGeeks ». Consulté le: 1 mars 2025. [En ligne]. Disponible sur: https://www.geeksforgeeks.org/catboost-ml/

[112] « What Is CatBoost? (Definition, How Does It Work?) », Built In. Consulté le: 1 mars 2025. [En ligne]. Disponible sur: https://builtin.com/machine-learning/catboost

[113] « The flow diagram of the CatBoost model », ResearchGate. Consulté le: 17 mai 2025. [En ligne]. Disponible sur: https://www.researchgate.net/figure/The-flow-diagram-of-the-CatBoost-model_fig3_370695897

[114] G. James, D. Witten, T. Hastie, et R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. in Springer Texts in Statistics. New York, NY: Springer US, 2021. doi: 10.1007/978-1-0716-1418-1.

[115] C. G. Misar, L. Xu, A. Boe, R. N. Gates, P. S. Johnson, et A. E. Olson, « Validation of a Technique for Estimating Alfalfa (Medicago sativa) Biomass from Canopy Volume », *Am. J. Plant Sci.*, vol. 7, nº 1, Art. nº 1, janv. 2016, doi: 10.4236/ajps.2016.71023.

[116] *Pattern Recognition and Machine Learning*. Consulté le: 6 avril 2025. [En ligne]. Disponible sur: https://link.springer.com/book/9780387310732

[117] A. Field, *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications, 2024.

[118] « Figure 1: Linear regression model. », ResearchGate. Consulté le: 17 mai 2025. [En ligne]. Disponible sur: https://www.researchgate.net/figure/Linear-regression-model_fig1_381857634