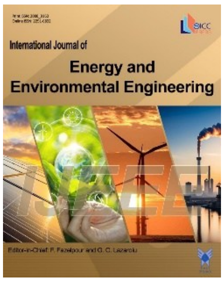

Volume 15, Issue 3, (2024) Article 11

International Journal
of
Energy and Environmental Engineering (IJEEE)

https://doi.org/10.57647/ijeee.2024.1503.11

Performance Analysis of Automatic Algorithm Learning Techniques for Wind Speed Prediction and Modeling: The Case of the City of Amdjarass

Osée Mounkang[1,2*], Claude Vidal Aloyem Kaze[3,4], Nzoko Tayo Dieudonné[1], Germaine Djuidje Kenmoe [1], René Tchinda[3]

[1] *Environmental Energy Technologies Laboratory (EETL), Department of Physics, University of Yaounde I, P.O Box 812, Yaounde, Cameroon*

[2] *Department of Physics, University of Pala, P.O Box 20, Pala, Chad*

[3] *Department of Energetic, Environment and Thermal Engineering, UR-ISIE, University Institute of Technology Fotso Victor, University of Dschang, P.O Box 134, Bandjoun, Cameroon*

[4] *HTTTC Higher Technical Teacher's Training College, Department of Renewable Energy, Kumba, Cameroon*

***Corresponding author:** osee.mounkang@facsciences-uy1.cm

Original Research

Abstract

Wind speed forecasting is essential for electrical energy management and planning. Wind turbine energy production varies according to fluctuations in wind speed. Unpredictable variations in wind patterns create vulnerabilities for wind power installations. To mitigate this unpredictability, an effective approach is to anticipate wind speed at specific heights, which is crucial for the operation of wind farms. Although previous work has examined different learning algorithms, researchers are still striving to find a stable model that minimizes prediction error. The objective of this study is to predict wind speed using machine learning algorithms. The results of the study highlight the remarkable predictive performance of eleven models, which are based on regression and classification methods with cross-validation. On the test data, the evaluation reveals impressive results, with coefficients of determination ranging from 0.986023 to 0.794080. The comparative study of the different algorithms shows that the proposed LightGBM (LGB) model outperforms most similar models in the literature, achieving statistically significant accuracy values. This result made it possible to evaluate the power of a wind turbine over the 12 months of the year. It shows that March generates the most energy, while August produces the least.

Keywords: Low-energy technologies, Geological routing, Educational projects, Energy efficiency, Sustainable development

Cite this article: Mounkang O., Kaze C.V.A., Dieudonné N.T., Djuidje Kenmoe G., Tchinda R., Performance Analysis of Automatic Algorithm Learning Techniques for Wind Speed Prediction and Modeling: The Case of the City of Amdjarass. Int. J. Energy Environ. Eng. 2024; 15(3) : 1-22 https://doi.org/10.57647/ijeee.2024.1503.11

Nomenclature

LR	LINEAR REGRESSION
LASSO	Least Absolute Shrinkage and Selection Operator
GBR	Gradient Boosting Regression
MLP	Multi-Layer Perceptron
KNN	K-Nearest Neighbors
RFR	Random Forest Regression
CART	Classification and Regression Trees
ABR	Adaptive Boosting Regression
ETR	Extra Trees Regression
XGB	(XGBoost, eXtreme Gradient Boosting
LGB	LightGBM, Light Gradient Boosting Machine
WED	Wind Energy Density
V	Wind speed, m/s
$f(v)$	Probability of observing wind seed
$p(v)$	Wind power, kwh
V_C	Cut-in wind speed, m/s
CF	Capacity factor, %
N	The number of data pairs
X	The values of the first variable
Y	The values of the second variable
$\sum xy \sum xy$	The sum of the values of the first variable
$\sum x \sum x$	The sum of the values of the first variable
$\sum y \sum y$	The sum of the values of the second variable
$\sum x^2 \sum x^2$	The sum of the squares of the values of the first variable
$\sum Y^2 \sum Y^2$	The sum of the squares of the values of the second variable
k	Shape factor
c	Scale factor (m/s)
z_1, z_2, z_3	Heights
$v(z)$	Reference speed
n	Total number of trees in the ensemble
$f(z)$	Activation function
w	Weight associated with each connection
$h(x)$	Specific regression model
λ	Bias or intercept of the model
$\hat{y}_i$	Predicted value
α	Pegularization parameter
β_j	Associated coefficient
$J(B)$	Cost function to minimize

1. Introduction

The increasing worldwide demand for energy, coupled with the steady decline of fossil fuel reserves, calls for innovative and swift solutions, especially in the traditional energy sector. Numerous nations have recognized the need to invest in sustainable energy sources to address present and future power requirements [1]. This shift has spurred advancements in technology, the creation of cutting-edge systems, and the execution of projects utilizing diverse renewable energy options. Among these alternatives, wind energy emerges as a highly efficient solution, owing to its cost-effectiveness and broad accessibility [2]. Wind speed remains a key parameter to assess, both during the planning stages and after a wind farm's commissioning [3].

Accurate wind speed analysis has become a key focus in scientific research, with extensive studies dedicated to understanding its behavior to enhance prediction reliability. Nevertheless, the unpredictable and highly variable nature of wind power poses substantial challenges for incorporating medium- to large-scale wind energy systems into existing power networks [4, 5]. These fluctuations can compromise grid stability and reduce operational efficiency, particularly when wind generation is integrated into the main electrical grid [6, 7].

Reliable wind speed forecasting is crucial for promoting equitable energy access and optimizing power generation planning and grid distribution [8]. Accurate predictions also help maintain cost competitiveness by reducing energy waste and enhancing operational efficiency [9]. However, wind speed's inherent variability and unpredictability create significant challenges for direct estimation, leading to uncertainties in wind turbine power output [10]. To tackle these challenges, researchers in energy and environmental sciences have developed and refined a wide range of wind speed prediction methods [11]. Today, multiple forecasting models exist, each employing distinct strategies and operating across different time horizons. These models are systematically classified using established frameworks from the literature [12]. The majority of forecasting techniques can be grouped into two primary approaches: (1) physics-based models, which incorporate terrain data, temperature, and wind farm geometry, often using numerical weather prediction (NWP) systems; and (2) data-driven statistical methods, which analyze historical wind patterns from specific locations [4, 13-15].

10.57647/ijeee.2024.1503.11

Physics-based approaches utilize environmental parameters including topography, thermal conditions, and turbine arrangement configurations, typically incorporating outputs from atmospheric modeling systems like numerical weather prediction (NWP) [16]. Data-driven techniques, conversely, process extended historical wind velocity records from localized regions, applying iterative algorithms optimized for near-term projections despite their intensive processing requirements [17,18]. Unlike their physics-based counterparts, these analytical models extract predictive relationships directly from measurement datasets without dependence on pre-specified physical formulations [19].

Our study uses machine learning methods to assess long-term wind patterns and predict wind speeds. This work was initiated in response to severe energy shortages affecting Amdjarass in eastern Chad, where conventional energy supplies have proven inadequate. In response to this situation, the Chadian authorities have recently placed an emphasis on green energy by providing targeted support for sustainable projects in the sector.

Research has identified machine learning models that can provide accurate wind speed predictions [20]. This study aims to enrich the knowledge base by proposing ideas that could improve wind energy production planning in the city of Amdjarass. The conclusions drawn from the study establish the supremacy of regression and classification models, not only in terms of accuracy but also in terms of computation time. The implications of this study extend beyond academic boundaries, providing valuable guidance to wind energy planners and decision-makers in Chad, as well as in provinces with similar climatic conditions [21].

Several comparable experimental investigations have been performed across different meteorological stations. Notably, Khosravi et al. [22] evaluated wind speed prediction accuracy in Iran using three distinct machine learning approaches: Support Vector Regression (SVR), an Adaptive Neuro-Fuzzy Inference System (ANFIS), and a Multilayer Perceptron Neural Network (MLP). Their study incorporated four key input variables: temporal data, atmospheric pressure, ambient temperature, and relative humidity. Comparative analysis revealed SVR's superior performance, demonstrating the lowest prediction error with a Mean Squared Error (MSE) of 0.378 m/s, Root Mean Squared Error (RMSE) of 0.6940 m/s, and an exceptional coefficient of determination (R^2) of 0.9704.

The study Démolli et al., [23] evaluates machine learning algorithms for wind speed prediction, highlighting their usefulness for current research. The best-performing models include XGBoost, which demonstrates strong long-term forecasting ability, and Support Vector Regression (SVR), which stands out when standard deviation is excluded from the input features (R^2 = 0.955, MAE = 32.63). Random Forest (RF) achieves the best overall performance with an R^2 of 0.995 and an MAE of 7.048, while LASSO shows limited effectiveness with an R^2 of 0.862. These results provide a useful benchmark for selecting wind speed prediction models in ongoing research.

The study conducted in Surat, India, evaluated eight machine learning algorithms for wind speed prediction, using hourly data from 2010 to 2019. The algorithms analyzed include linear regression, Gradient Boosting, AdaBoost, decision tree, Random Forest, LSTM networks, Multilayer Perceptron, and RNN. Linear regression proved particularly effective, with a very low MAE of 0.310 and a coefficient of determination (R^2) of 0.987, highlighting its potential for accurate forecasting [24].

The study examined four machine learning models for predicting wind speed and their impact on wind energy production in Bangladesh. The models tested were the Decision Tree Regressor, the Gradient Boosting Regressor, the Random Forest Regressor, and the Voting Regressor. The Gradient Boosting Regressor showed promising results, with a coefficient of determination (R^2) of 0.8621 for Kutubdia Island, indicating a good correlation with wind energy production.

Aman et al. [25] predicted very short-term wind speed in Canada using four machine learning algorithms: the Multilayer Perceptron Regressor, the Decision Tree Regressor, the K-Nearest Neighbors Regressor, and the Random Forest Regressor. They found that the Multilayer Perceptron Regressor provided the best prediction accuracy, with an R^2 of 95.3%.

Thus, the objective of this work is to predict wind speed in Amdjarass using machine learning algorithms. Despite the availability of similar studies, the literature emphasizes that wind forecasting depends on the site. This means that the optimal forecasting models for one location may not be the most suitable for others. This analysis describes the wind conditions in the region and provides valuable feedback for investors in this part of Chad.

The data used in this study were collected at the N'Djamena weather station department, using the station installed in Amdjarass. The data cover a period from January 1, 2015, to December 31, 2023 (eleven years, or a total of 78,887 values recorded at one-hour intervals). The models used in this study are as follows: the Linear Regression (LR) algorithm, the Least Absolute Shrinkage and Selection Operator (LASSO), the K-

Neighbors Regressor (KNN), the Decision Tree Regressor (CART), the Multi-Layer Perceptron (MLP), the AdaBoost Regressor (ABR), the Gradient Boosting Regressor (GBR), the Random Forest Regressor (RFR), Extra Trees Regressor (ETR), LightGBM (LGB), and the LightGBM (LGB) in the context of wind speed prediction for the city of Amdjarass.

In a study conducted in Romania, a comparison was made between four algorithms Artificial Neural Networks (ANN), Support Vector Regression (SVR), Random Forests, and Random Trees for predicting wind speed. The authors concluded that SVR offered the best accuracy. In a similar study, Wang, T. [26] proposed a combined model for predicting short-term wind speed based on empirical model decomposition, feature selection, SVR, and LASSO with cross-validation. The dataset was collected from two wind stations located in Michigan, USA, and the results demonstrated that the combined model effectively predicted wind speed.

2. Materials and methods

2.1. Presentation of the study site

Located between the Tropic of Cancer and the 8th parallel north, Chad has an area of 1,284,000 km² [27]. It possesses wind potential with wind speeds ranging between 1.9 and 25 m/s, which constitutes an interesting energy resource for the country. The city of Amdjarass is located in eastern Chad, in the province of Enedy East, on the border with northern Sudan. Rainfall is rare in Amdjarass, and the few occurrences of rain are often accompanied by strong winds, which can exceed 25 m/s. Temperatures fluctuate between 20 and 45° Celsius during the day. Nights can be cool, with temperatures dropping to minus 10 degrees Celsius. The weather can become windy, with sandstorms [28]. It is important to note that the climate in desert regions such as Amdjarass is generally characterized by significant variability from year to year in terms of precipitation. Periods of drought can have a major impact on water availability and the daily lives of the province's inhabitants. In the town of Amdjarass, there is a power plant consisting of four Vergnet wind turbines capable of delivering 1,100 kW of power, supplying this part of Chad with energy. The study focuses on the town of Amdjarass, located at 15.9722° north latitude and 22.7702° east longitude, and Figure 1 shows a representation of our study site.

2.2. Machine learning algorithms

Machine learning (ML) enables systems to understand on their own and then evaluate certain outputs [28, 29]. The performance of a machine learning algorithm depends on the selection of attributes and the success of the learning process [30]. In this research, eleven machine learning algorithms were applied. Figure 2 shows the different machine learning algorithms used.
These are the LR, LASSO, KNN, CART, MLP, ABR, GBR, RFR, ETR, XGB and LGB models. The various algorithms were executed using Python version 3.9.12, within a Jupyter Notebook environment. A total of eleven machine learning algorithms were implemented. It should be noted that the performance of a machine learning algorithm is analyzed based on the accuracy obtained and the time required to build the model. Building a model involves developing a probabilistic model that best describes the relationship between the dependent and independent variables. The complete methodology used for this research is fully presented in Figure 3.

The section below describes the various stages of data collection, pre-processing, application of machine learning algorithms, and wind speed prediction.

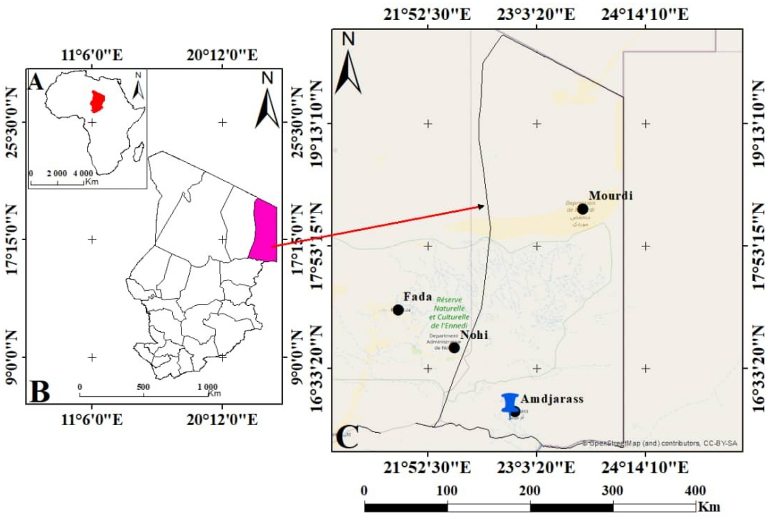

Figure 1. Map of the city of Amdjarass

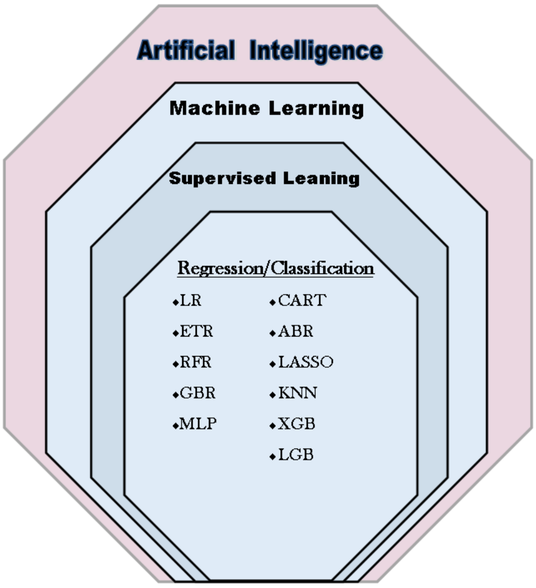

Figure 2. Structure of the study base

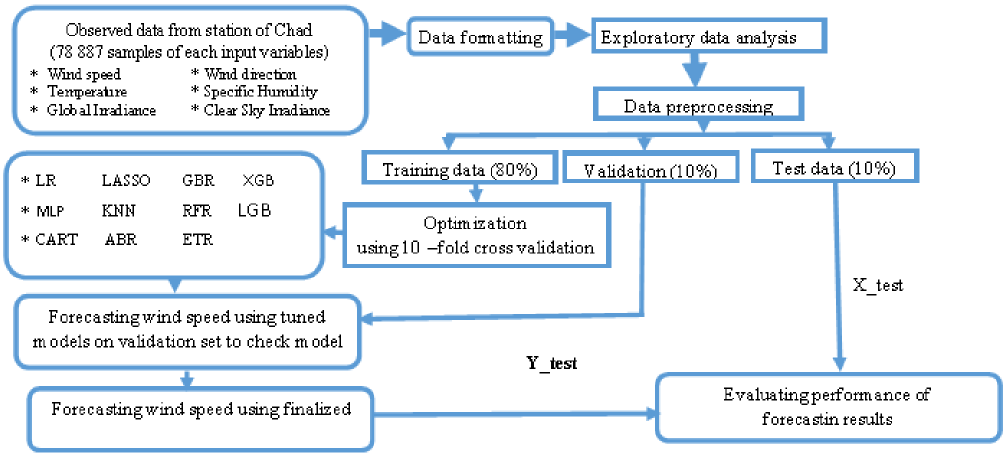

Figure 3. Preliminary phase and data learning

2.3. Data collection

Step 1. Data collection and formatting: Initially, observed data (wind speed, wind direction, temperature, relative humidity, global solar radiation, and clear-sky radiation) from January 1, 2015, to December 31, 2023 (78,887 data samples over eleven years) were collected. Data formatting was performed by removing irrelevant information, taking into account the correlation calculations performed beforehand, and reorganizing the necessary parts.

Step 2: Exploratory Data Analysis (EDA) serves as a critical preliminary phase in machine learning pipelines, enabling comprehensive examination of dataset characteristics and inherent patterns. This stage involves systematic investigation of descriptive statistics and data distributions to uncover meaningful insights from the processed information. Through rigorous EDA, researchers can identify trends, anomalies, and relationships within the data that inform subsequent modeling decisions.

Step 3: Data Preprocessing: The processing of input variables for wind speed forecasting is a crucial step in preparing data so that it is suitable for use in forecasting models. After data collection, processing involves removing outliers based on statistical analysis of the data, replacing them with the mean or median, and using predictive models to estimate missing values. In general, the processing of recording errors, outliers, and missing values aims to ensure the quality and reliability of the data before it is used in analyses or predictive models **.** This improves the accuracy of the results and enables informed decisions to be made based on reliable data.

Step 4. Train-test split: After preprocessing, the dataset is divided into three subsets: i) the training set (80%), ii) the validation set (10%), and iii) the test set (10%).

Step 5. Model optimization and training: At this stage, eleven (09) machine learning algorithms are deployed to predict wind speed. For algorithm optimization, k-fold cross-validation is implemented, with and without parameter tuning.

Step 6: The trained algorithms undergo rigorous evaluation by comparing their performance on the validation dataset against the baseline model developed from the training data. When performance discrepancies fall within acceptable thresholds, the models are subsequently assessed on the test dataset, where predicted outputs are benchmarked against ground truth observations to validate system accuracy. This study conducted an exhaustive comparative analysis between measured and predicted wind speeds, highlighting the practical advantages of advanced forecasting methodologies. Machine learning regression models provide a diverse toolkit for wind speed prediction, with each algorithm offering distinct advantages tailored to particular application scenarios. Optimal model selection requires careful consideration of multiple factors, including: Dataset properties and quality, available computational resources, specific forecasting objectives and precision requirements.

2.4. Optimization and fine-tuning of models: K-fold cross-validation

Optimization and hyperparameter tuning can significantly improve the performance of regression models and their ability to generalize to new data. A common method for determining the performance of an ML model is k-fold cross-validation**.** The training data is divided into k-folds**,** or subsets, to apply this technique. The model is then trained and evaluated multiple times on these folds. Each fold serves as the validation set once, while the remaining folds are used for training.

2.5. Choosing a wind turbine

The power output of a wind turbine varies according to the effect of wind on the rotor. The power curve represents the power produced as a function of wind speed. The choice of a wind turbine must be based on criteria such as start-up speed, mast height, maximum power output, and air density. The characteristics of the GEV MP 275/32 wind turbine model were selected for this study [31]. These characteristics are presented in Table 2.

Based on the characteristics of our wind turbine in the table, the power curve shown in Figure 4 can be observed.

The wind turbine begins to generate energy at its start-up speed of 3.5 m/s. Maximum power is reached when the wind speed reaches 10 m/s. This power remains constant as the speed increases up to 25 m/s. If the wind speed exceeds 25 m/s, the wind turbine ceases operating.

2.6. Data preprocessing

Here are some ML steps for preprocessing data.

Step 1. A library is a collection of modules that can be called and used. In Python, numerous libraries are called for preprocessing, namely NumPy, Matplotlib, Pandas, Scikit-learn, and Seaborn.

Step 2. Load the collected data. The data is available in Excel .xlsx format. To load an .xlsx file, use the read_excel.xlsx function from the panda's library. Once the dataset has been loaded, they inspect it and search for any missing data in the database. They create a matrix of characteristics X and an observation vector Y relative to X;

Step 3. Checking for missing values. Once the feature matrix has been created, they look for missing values.

Step 4. Verification of categorical data. The data in the dataset must be in numerical form in order to perform calculations.

Step 5. Feature scaling. Feature scaling is a technique used to bring data values into a shorter range.

Step 6. Data division. The values are divided into different subsets: 80% for training to build the model, 10% for validation, and 10% of the remaining data set for testing the models on generic data.

Table 1. Characteristics of the GEV MP 275/32 wind turbine

Features	Specifications
Startup speed	3.5 m/s
Stopping speed	25 m/s
Height of mast	55-67 m
Rotor diameter	32 m
Maximum power	275 KW
Density	1.025 Kg/m^3
Manufacturer	Bonfiglioli, CMD, Winergy
Swept area	805 m^2

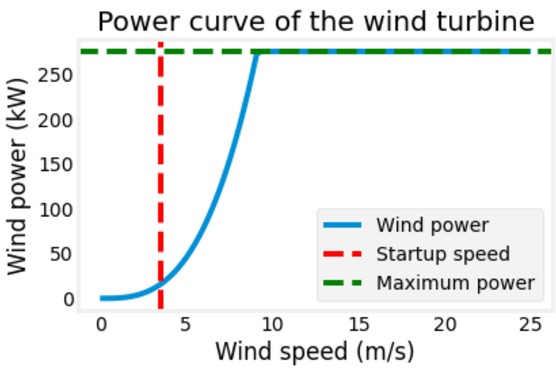

Figure 4. Power curve of the GEV MP 275/32 wind turbine

10.57647/ijeee.2024.1503.11

The transformation and loading process plays a key role in ensuring the integrity and reliability of all data used for wind speed prediction. By addressing data quality issues, transforming skewed distributions, and removing outliers, the dataset provides meaningful information for subsequent model analysis according to Smith et al., 2019 [11].

2.7. Correlation calculation

This subsection aims to determine the input variables most correlated with the output variable. To do this, the correlation coefficient is calculated using formula 1 [32-34].

$$R = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

This correlation coefficient calculation made it possible to identify the different variables that influence wind speed. It should be noted that a correlation coefficient value closer to 1 implies a better forecast. This indicator varies between -1 and 1. When the correlation coefficient gives a value of 1, this symbolizes a perfect agreement between the measured value and the output variable, and a value of -1 indicates a perfect opposite agreement. A value of 0 means that the input variable has no influence on the output value.

2.8. Machine learning algorithms

This study employs classical machine learning and deep learning algorithms for solar radiation forecasting. The classical approaches include:

Linear Regression (LR) [35], models linear relationships between input features and solar radiation;

LASSO Regression [35], extends LR with L1 regularization for feature selection;

K-Nearest Neighbors (KNN) [21, 22], predicts based on average values of k-nearest data points;

Decision Trees (CART) [20], uses hierarchical feature splits for prediction;

MLP Regressor [20], neural network with hidden layers for non-linear modeling;

AdaBoost (ABR) [35], sequentially combines weak learners with weight adjustments;

Gradient Boosting (GBR) [18,36], iteratively builds trees to correct previous errors;

Random Forest (RFR) [37], ensemble of decision trees with random feature subsets;

Extra Trees (ETR) [20, 38, 39], similar to RFR but with random split thresholds;

XGBoost (XGB) [40], an optimized implementation of Gradient Boosting with built-in regularization and missing value handling, renowned for its high performance in machine learning competitions.

LightGBM (LGB) [41], an alternative to XGBoost that uses asymmetric tree growth and histograms to speed up training, particularly effective for large data volumes and real-time applications.

2.9. Evaluation measures

To compare the performance of eleven machine learning models, the hold-out and k-fold cross-validation techniques were used to evaluate the performance of each model. These different statistical parameters are summarized in Table 2.

2.10. Weibull adjustment on wind speed distribution

Before using Weibull parameters to model wind speed, it is a good idea to check whether this distribution is applicable to our site [48]. We therefore perform a chi-square goodness-of-fit test. This involves plotting the scatter plot of the wind speeds measured at the site as a function of the Weibull parameter. The test is applicable when wind speeds vary in the range of 1 to 30 m/s. This distribution is shown in Figure 5.

Figure 5 shows the characteristics of the chi-square distribution curve. The curve has an asymmetrical shape with an elongated right tail, which means that the degree of freedom is high, as it approaches a normal distribution.

2.11. Modeling wind speed using the Weibull distribution

The Weibull distribution, characterized by its scale (λ) and shape (k) parameters, is a preferred tool for analyzing wind speeds and estimating wind turbine performance.

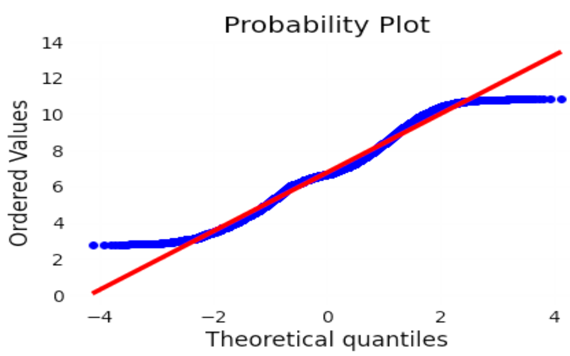

Figure 5. Weibull compatibility using the chi-square test

Table 2. Summary of statistical measures for evaluation

Metrics	Equation	Description
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y - \hat{y}_i \rvert$ (11)	The mean absolute error is a quantity often used to measure the difference between observed values and predicted values. Its mathematical formula is given by equation (11) [42].
MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ (12)	The mean square error allows several estimators to be compared, particularly when one of them is biased; it is given by equation (12) [43].
RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$ (13)	RMSE is a measure of the variation in predicted values around measured values. The smaller its value, the better the model [44]. The square root of the mean square error is defined according to formula (13) [45].
nRMSE	$nRMSE = \frac{RMSE}{\bar{y}}$ (14)	The Normalized RMSE scales RMSE by the mean of observed values $\bar{y}$, enabling comparison across different scales or units. Values closer to 0 indicate better fit [46].
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2}$ (15)	The coefficient of determination R^2 is a statistical measure that shows how well a model's predictions correspond to actual values [46, 47]. It is defined by formula (15).
MAPE	$MAPE = \sum_{i=1}^{n}\left\lVert\frac{y_i - \bar{y}}{y_i}\right\rVert \times 100$ (16)	The Mean Absolute Percentage Error expresses errors as a percentage of actual values, making it scale-independent. Useful for comparing models across datasets [44]. *Limitation*: Undefined for zero values in y_i. It is defined by formula (16).

The probability density function (Eq. 17) and cumulative distribution function (Eq. 18) describe the variability of wind speeds, while the average speed is given by Eq. 19 [9,49].

$$f(v) = \frac{k}{c} \times \left(\frac{v}{c}\right)^{k-1} exp\left[-\left(\frac{v}{v}\right)^{k}\right] \quad (17)$$

$$f(v) = \int_0^v f(v)dv = 1 - exp\left[-\left(\frac{v}{c}\right)^{k}\right] \quad (18)$$

$$v_m = \int_0^\infty v \times f(v)dv \quad (19)$$

The parameters k and c are determined using standard methods (graphical, moment, maximum likelihood). Variability is quantified using Eq. 20, with the scale factor c calculated using Eq. 21. The shape of the distribution and its asymmetry are characterized by Eqs. 22-23 [9,50].

$$k = \begin{cases} 1.05 \times v^{0.5} if v \prec 3 \\ 0.94 \times v^{0.5} if 3 \prec v \prec 4 \\ 1.05 \times v^{0.5} if v \succ 4 \end{cases} \quad (20)$$

$$c = \frac{v}{\Gamma\left(1+\frac{1}{k}\right)} \quad (21)$$

$$\Gamma(x) = \left(\sqrt{2\pi x}\right)(X^{X-1})(e^{-X})\left(1 + \frac{1}{12}X + \frac{1}{288}X^2 - \frac{139}{51840}X^3 + \ldots\right) \quad (22)$$

$$X = 1 + \frac{3}{k} \quad (23)$$

For wind turbine installation, we extrapolate the measurements taken at a height of 10 m to operational heights (55 m, 67 m) using the power law (Eqs. 24-25) [51].

The corresponding adjustments to the Weibull parameters follow Eqs. 26-28 [7,10].

$$v(z_1) = v(z_0) \times \left(\frac{z_1}{z_0}\right)^{\alpha} \quad (24)$$

$$\alpha = \frac{0.37 - 0.088\, ln(v_{z_0})}{1 - 0.00881\, ln\left(\frac{Z_0}{10}\right)} \quad (25)$$

$$n = (0.37 - 0.088\, ln(C_{10})) \quad (26)$$

$$k_z = \frac{k_{z_0}}{1 - 0.00881\, ln(z/10)} \quad (27)$$

$$C_z = C_{z_0} \times (z/z_{10})^n \quad (28)$$

Wind energy density (Eq. 29) and recoverable power (Eqs. 30-33) can be used to estimate the site's potential. The choice of wind turbines takes into account the load factor (LF, Eq. 34), which incorporates site-specific parameters (k, c) and the technical characteristics of the wind turbines (VD, VN, VA) [6,12]. Sites with LF > 25% are considered viable for energy production [52].

$$WPD = p(v) = \frac{P(v)}{S} = \frac{1}{2}\rho \times c^3 \times \Gamma(x) \quad (29)$$

$$WED = p(v)T = \frac{P(v)}{S}T = \frac{1}{2}\rho \times v^3 \times \Gamma(x)T \quad (30)$$

$$P_u = \begin{cases} 0(v < v_{start}) \\ \frac{1}{2}\rho c^3 \Gamma(x)(v_{start} \le v \le v_{rat}) \\ P_n(v_{rat} \le v \le v_{stop}) \\ 0(v_{stop} < v) \end{cases} \quad (31)$$

$$P_u = \int_0^{\infty\int} \frac{1}{2}\rho c^3 \Gamma(x) f(v,k;c)dv + \int_{v_{start}}^{v_{rate}\int} P_n f(v,k,c)\,dv \quad (32)$$

$$P_u = P_n \times FC \quad (33)$$

The load factor of a wind turbine is an indicator that measures the efficiency and performance of the wind turbine in converting wind energy into electrical energy. It is expressed as a percentage and represents the amount of energy actually produced by the wind turbine in relation to its theoretical maximum capacity. Five parameters are taken into account when calculating the load factor (LF). Two depend on the chosen site (the shape factor k and scale factor c), and three are provided by the wind turbine manufacturer (start-up speed VD, nominal speed VN, and wind turbine cut-out speed VA) [6, 12].

$$FC = \frac{exp\left(-\left(\frac{V_D}{C}\right)^K\right) - exp\left(-\left(\frac{V_N}{C}\right)^K\right)}{\left(-\left(\frac{V_N}{C}\right)^K\right) - \left(-\left(\frac{V_D}{C}\right)^K\right)} - \left(exp - \left(\frac{V_A}{C}\right)^K\right) \quad (34)$$

When the load factor exceeds 25%, it can be concluded that this site can produce electrical energy thanks to wind speed [53].

3. Results and discussions

3.1. Data exploration and processing

Table 3 presents the descriptive analysis of all variables used. It includes the total value, mean, standard deviation, as well as minimum, maximum, and percentile values.

Table 3 shows that there are 78,887 recorded values. The average wind speed is 6.90 m/s with a standard deviation of 2.96. The percentiles are 4.52, 7.17, and 9.44 m/s, respectively, while the maximum wind speed recorded at 10 m above ground level is 17.81 m/s. Air temperatures vary between 1.23 and 44.19 °C, with an average value of 26.54 °C. The percentiles are 21.51, 26.65, and 32.30 °C. Finally, relative humidity values range from 0.55 to 20.08%, with an average of 5.91% and a standard deviation of 4.47%. The percentiles are 2.75, 3.97, and 8.00%, respectively. Solar irradiance on a clear day varies between 0.01 and 14.92 W/m², with an average of 4.95 W/m² and a standard deviation of 2.27. The percentiles are 3.19, 4.87, and 6.55 W/m², respectively. Total irradiance fluctuates between 0.00 and 1128.92 Wh/m², with an average of 281.10 Wh/m² and a standard deviation of 360.90. The percentiles are 0.00, 19.23, and 612.30 Wh/m², respectively.

3.1.1. Calculation of correlation between different variables

Table 4 shows the correlation values obtained from the different variables used to predict wind speed.

We observe that the correlation value varies depending on each variable. Clear- sky irradiation has a value of 0.88, the highest among the other variables. This indicates that an increase in the input variable is associated with an increase in the output variable, demonstrating a positive relationship. The correlation between temperature and wind speed is -0.57, indicating a moderate relationship in the negative direction. This means that the two variables move in opposite directions. A similar trend is also observed between global radiation, wind direction, and relative humidity. The correlation value of 1 presented in Table 4 is expected, as it indicates a perfect correlation between a variable and itself, which is normal. Figure 6 shows the autocorrelation coefficient as a function of time lag ranging from 0 to 70,000 units. Autocorrelation values range from -1.00 to +1.00, reflecting the intensity and direction of temporal dependence in the data. For short time lags, the coefficients are mostly positive, indicating strong persistence in the data. As the time lag increases, the autocorrelation decreases and approaches zero. This suggests that the dependence between observations weakens over time, eventually becoming negligible. For some lags, the coefficients become negative, revealing an inverse relationship.

3.1.2. Viewing variables

Figure 7 illustrates the chronological distribution of the different variables as a function of wind speed. These scatter plots allow us to visualize the distribution of all the measurements used in our study.

3.1.3 Wind speed extrapolation

Wind speed is proportional to height, as shown in Figure 8, which illustrates the change in speed at 10 m, 55 m, and 67 m above ground level. It can be seen that wind speed is higher at 55 m and 67 m.

The extrapolation of wind speed based on the Weibull parameters is also shown in this figure 8. It can be seen that at 55 m, speeds are higher than those measured at 10 m, while at 67 m, speeds exceed those at 55 m. The minimum speed at 55 m above the ground is 5.01 m/s, with a maximum value of 9.97 m/s. In contrast, the minimum speed at 67 m is 5.26 m/s, and the maximum value reaches 10.47 m/s.

3.2. Prediction with Machine Learning Algorithms

3.2.1. Evaluation of R^2, RMSE, nRMSE, MAPE and MAE Scores

This research aims to predict wind speed in the city of Amdjarass using eleven (09) machine learning algorithms. To evaluate the performance of these algorithms, three statistical measures commonly used in the literature are examined. Table 5 presents the results of these three statistical measures to identify the best prediction model. It also shows the different numerical values of the metrics calculated to evaluate the algorithms in the study. In this section, the examined algorithms will be analyzed based on the data provided in Table 5.

The K-fold cross-validation approach is applied during the evaluation process. The R-squared (R^2), RMSE (Root Mean Square Error), MAE (Mean Absolute Error) scores, and computation time are the main evaluation metrics.

These measures provide an objective assessment of each model's ability to reliably and effectively predict wind speed, offering insights into the advantages and disadvantages of each model in identifying underlying patterns in wind speed data. Based on the results in Table 5, each model used in our study will be analyzed in detail.

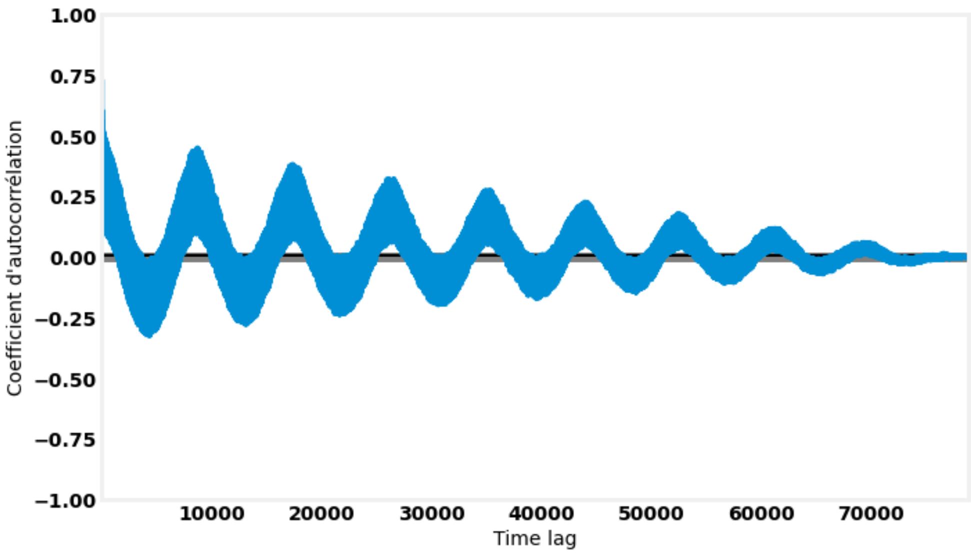

Figure 6. Autocorrelation curve

10.57647/ijeee.2024.1503.11

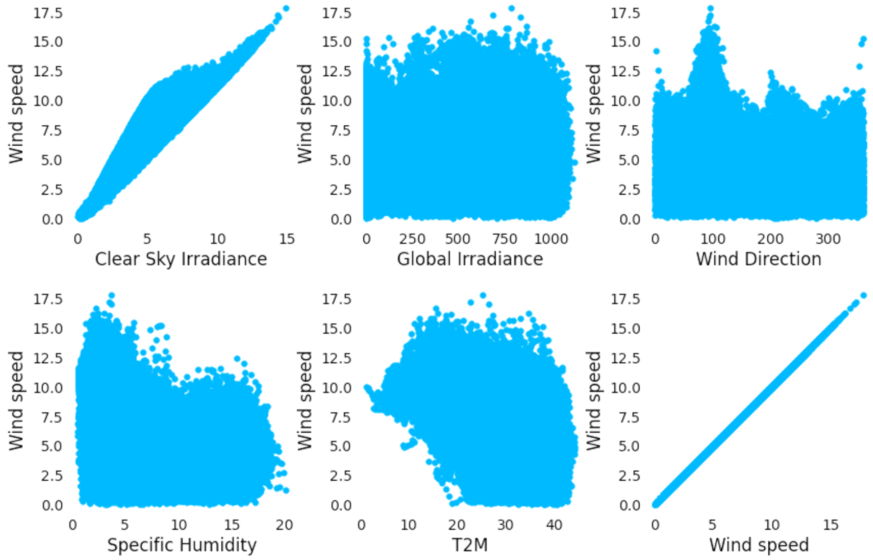

Figure 7. Visualization of individual variables as a function of wind speed

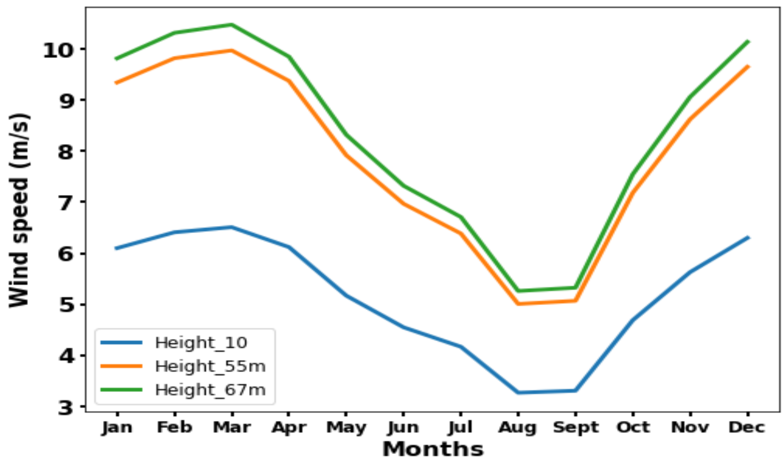

Figure 8. Wind speed curve at different heights

Table 3. Static description of the different values

Variables	Count	Mean	Std	Min	25%	50%	75%	Max
Global Irradiance	78887	281.1056	360.90429	0.00	0.00	19.23	612.30	1128.4
Clear Sky Irradiance	78887	4.955154	2.273193	0.01	3.19	4.87	6.55	14.92
Wind Direction	78887	132.3698	81.25037	0.00	82.98	95.69	173.57	359.95
Wind speed	78887	6.901913	2.966744	0.05	4.52	7.17	9.44	17.81
Specific Humidity	78887	5.915409	4.471794	0.55	2.75	3.97	8.00	20.08
Temperature	78887	26.5488	7.760507	1.23	21.51	26.65	32.30	44.19

Table 4. Correlation of variables based on wind speed

Variables	Correlation values
Global Irradiance	-0.259584
Clear Sky Irradiance	0.880829
Wind Direction	-0.410074
Wind speed	1.000000
Specific Humidity	-0.451847
Temperature	-0.573909

Table 5. Statistical error measures for machine learning algorithms

Models	RMSE	MAE	R^2	MAPE	nRMSE	Time(s)
LGB	0.355507	0.018194	0.986023	4.8741	0.055	17.967
XGB	0.358564	0.017016	0.985782	4.7925	0.055	13.9419
ETR	0.352242	0.016888	0.986098	4.5907	0.055	215.27
RFR	0.353954	0.016953	0.985963	4.5891	0.056	371.03
GBR	0.410781	0.015177	0.981094	6.6412	0.065	122.27
MLP	0.432061	0.059048	0.979084	6.2716	0.115	441.20
CART	0.491286	0.014935	0.972957	15.999	0.077	8.40
ABR	0.702888	0.012625	0.944645	16.103	0.114	8.40
LR	0.851735	0.026800	0.918718	18.535	0.130	0.55
LASSO	1.028588	0.022151	0.881460	24.556	0.157	0.58
KNN	1.355681	0.019775	0.794080	34.126	0.210	9.22

LGB (LightGBM): The LGB model takes 17.967 seconds to run, which is quite fast compared to other models. It has an RMSE of 0.3555, indicating that the predictions of this model have an average error of 0.3555 m/s compared to the actual values. This RMSE is considered excellent. The MAE is 0.0182, indicating very little dispersion in the model's performance across different parts of the data, thus showing that the model is stable and generalizes well to unknown data. The MAPE is 4.8741, meaning that the average percentage error is 4.87%, indicating good relative accuracy. With an R^2 of 0.9860, the LGB model explains approximately 98.60% of the variance in the data. We conclude that this model is capable of explaining a large portion of the variability in the data and is well-fitted.

XGB (XGBoost): The XGB model takes 13.9419 seconds to calculate, which is also relatively fast. Its RMSE is 0.3586, indicating an average error of 0.3586 m/s compared to the actual values, which is very good. The MAE is 0.0170, showing low performance dispersion, indicating good stability. The MAPE is 4.7925, indicating an average percentage error of 4.79%. With an R^2 of 0.9858, this model explains approximately 98.58% of the variance in the data, showing that it is well-fitted and capable of capturing the variability in the data.

The ETR model takes 215.27 seconds to run, which is considerably longer than previous models. Its RMSE is 0.3522, representing an average error of 0.3522 m/s, which is an excellent result. The MAE is 0.0169, indicating very low dispersion, which demonstrates high stability. The MAPE is 4.5907, meaning that the average percentage error is 4.59%. With an R^2 of 0.9861, ETR explains approximately 98.61% of the variance in the data, making it a very high-performing model despite its long execution time.

The RFR model takes 371.03 seconds to run, which is the longest among the models analyzed. Its RMSE is 0.3539, indicating an average error of 0.3539 m/s. The MAE is 0.0170, showing low performance dispersion. The MAPE is 5.636, indicating an average percentage error of 5.64%. The R^2 is 0.9860, meaning that the model explains approximately 98.60% of the variance in the data. Although the execution time is high, its performance is comparable to that of the best-performing models.

The GBR model has an RMSE of 0.4108, indicating an average error of 0.4108 m/s, which is lower than that of previous models. The MAE is 0.0152, showing low performance dispersion. The MAPE is 6.6412, indicating an average percentage error of 6.64%. With an R^2 of 0.9811, the model explains approximately 98.11% of the variance in the data. Although its performance is slightly lower, it remains robust and reliable.

The MLP model takes 441.20 seconds to run, which is relatively long. Its RMSE is 0.4321, indicating an average error of 0.4321 m/s. The MAE is 0.0590, indicating some dispersion in performance. The MAPE is 11.585, meaning an average percentage error of 11.59%. With an R^2 of 0.9791, the model explains approximately 97.91% of the variance in the data. Although it is able to explain a large portion of the variability, its long execution time can be a disadvantage.

The CART model has an RMSE of 0.4913, indicating an average error of 0.4913 m/s. The MAE is 0.0149, showing low dispersion, but the R^2 is 0.9730, meaning that it explains approximately 97.30% of the variance in the data. The MAPE is 15.999, indicating an average percentage error of 15.99%. Its execution time of 8.40 seconds is fast, but its performance is inferior to that of the previous models.

The ABR model has an RMSE of 0.7029, indicating an average error of 0.7029 m/s, which is higher than the other models. The MAE is 0.0126, showing low dispersion, but the R^2 is 0.9446, meaning that it explains approximately 94.46% of the variance. The MAPE is 16.103, indicating an average percentage error of 16.10%. Although fast, its performance is less satisfactory.

The LR model has an RMSE of 0.8517, indicating an average error of 0.8517 m/s. The MAE is 0.0268, showing some dispersion. The MAPE is 18.535, meaning an average percentage error of 18.54%. With an R^2 of 0.9187, it explains approximately 91.87% of the variance in the data. Although it is very fast (0.55 seconds), its performance is fairly limited.

The LASSO model gives an RMSE of 1.0286, indicating an average error of 1.0286 m/s, which is high. The MAE is 0.0221, showing moderate dispersion. The MAPE is 24.556, meaning an average percentage error of 24.56%. With an R^2 of 0.8815, it explains approximately 88.15% of the variance. Although it is fast (0.58 seconds), its performance is inferior to that of the other models.

The KNN model has an RMSE of 1.3557, with an average error of 1.3557 m/s, indicating poor performance. The MAE is 0.0198, showing some dispersion. The MAPE is 34.126, which means an average percentage error of 34.13%. The R^2 is 0.7941, meaning that it explains approximately 79.41% of the variance in the data. Although its execution time is reasonable (9.22 seconds), its performance is the lowest among all the models analyzed.

prediction curve for the various models

The prediction curve shows the results of different machine learning models (LR, LASSO, KNN, CART, MLP, ABR, GBR, RFR, ETR, XGB, LGB) compared to actual values over a period of up to 16,000 hours, with a zoom on the 100-to-200-hour interval. Predictions vary between 0 and 17.5, showing significant dispersion in performance across models. Some algorithms, such as XGB or LGB, seem to better capture trends in actual data, while others show more pronounced deviations. Zooming in reveals fine details, where predictions converge or diverge from observed values, highlighting the importance of model choice for short- and long-term accuracy. This analysis highlights the complexity of temporal modeling and the need to optimize hyperparameters to improve performance.

Figure 10 shows the performance of eleven different models evaluated using three key indicators: the R^2 test, the RMSE test, and the nRMSE test. The scores, ranging from 0.00 to 1.25, allow for a comparison of the models' predictive accuracy and quality. A high R^2 score indicates a better fit to the data, while low RMSE and nRMSE values reflect smaller prediction errors. The visual comparison of bars or points for each model highlights performance differences, helping to identify the most effective approaches for the given task. This analysis is essential for selecting the optimal model based on the required precision and robustness criteria.

The figure 11 shows the MAPE test results for eleven different models, expressed as a percentage. MAPE values range from 4.59% to 34.13%, indicating a decline in model performance. The first five models show relatively low errors (between 4.59% and 6.27%), suggesting good accuracy, while the last four show significantly higher errors (up to 34.13%), indicating poorer fit to the data. These results highlight significant differences in performance between the tested models.

To evaluate the performance of the model, a scatter plot was used to compare the predicted wind speeds with the actual speeds from the test dataset.

Figure 12 illustrates the scatter plots representing the predicted wind speed values based on the observed values. This allows us to observe the graphical representation and examine the relationship between the predicted wind speed and the observed speed, as well as to evaluate the robustness of this association using the coefficient of determination.

Figure 13 shows the best results from eleven models used for speed prediction. This LGB (Light Gradient Boosting) method indicates a coefficient of determination R^2 of 0.98, indicating an excellent fit between the predictions and the actual values. The graphs show a comparison between predicted and actual values, as well as a zoom on these predictions for a more detailed analysis. The strong correlation suggests that the model accurately captures variations in wind speed over time, which is essential for applications such as weather forecasting or wind energy management. The horizontal axis represents time in hours, allowing the evolution of predictions over a specific period to be visualized. The predicted values generally follow the trend of the observed values, with few significant deviations between the two sets of values. This indicates that the model is well suited and that the predictions provided are highly accurate.

Figure 14 shows the Training Loss and Validation Loss curves of the LGB model for wind speed prediction. A gradual and convergent decrease in both curves indicates that the model is learning effectively without overfitting. The training loss and validation loss remain close, confirming that the model generalizes well. This analysis is crucial for validating the robustness of the model in real-world applications such as wind forecasting. The same is true for other models, which show no overfitting.

3.2.4. Presentation Confidence interval

The Figure 15 compares the performance of eleven machine learning models in terms of average RMSE and confidence intervals. The XGB, LGB, and GBR methods perform best with the lowest average RMSEs (0.4–0.6), while the LR, LASSO, and KNN models perform worst (0.8–1.4).

The accuracy of the estimates and the significance of the differences between models are noteworthy. This analysis confirms the superiority of the XGB and LGB approaches for complex predictive tasks, while highlighting the potential need for optimization for less performant models such as KNN or CART. The results are consistent with current knowledge in machine learning, where XGBoost and LightGBM are recognized for their effectiveness.

3.2.5. Similarities and consistencies

Numerous previous studies have explored wind speed prediction using machine learning models, with varying degrees of success.

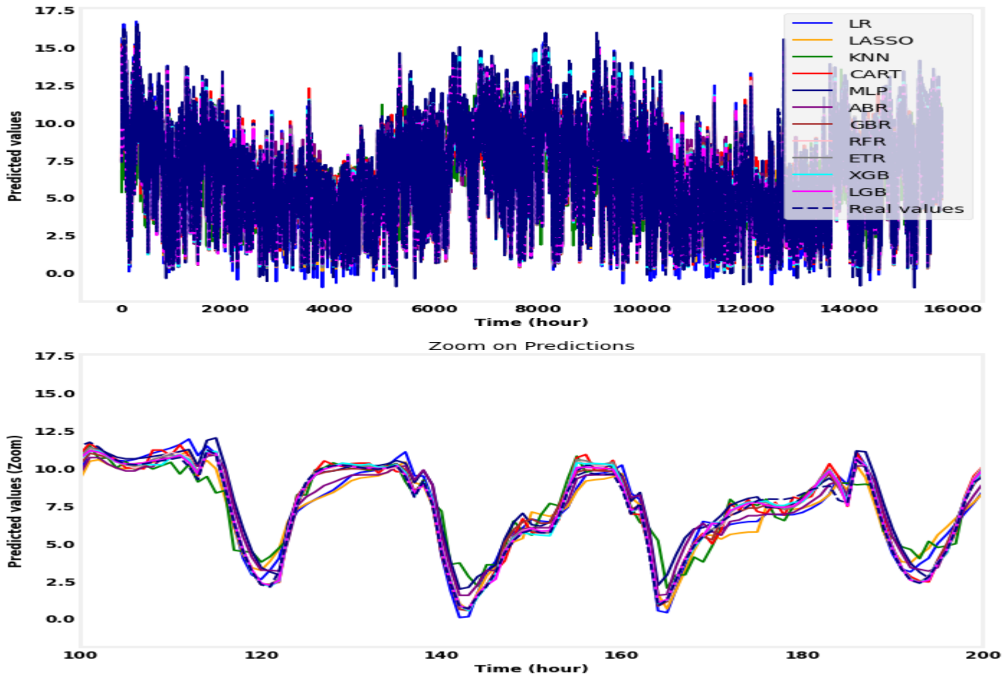

Figure 9. Comparative prediction curve of the different models

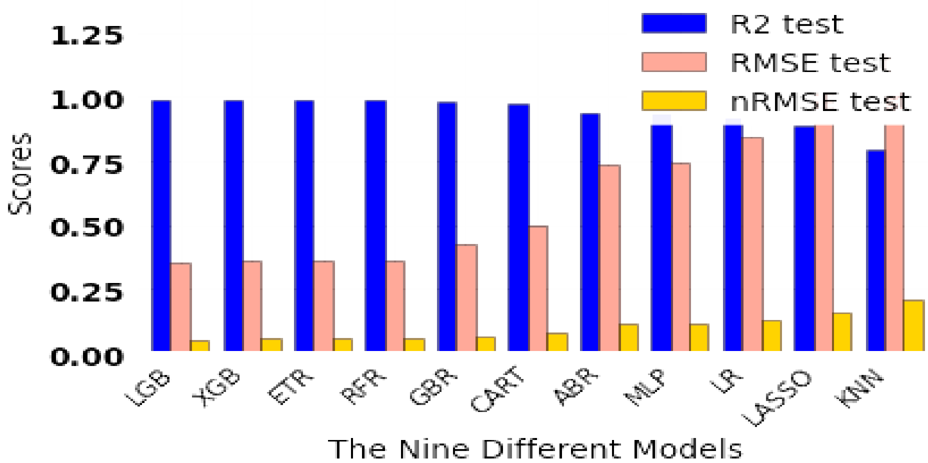

Figure 10. Comparative performance (R^2, nRMSE and RMSE) of the models.

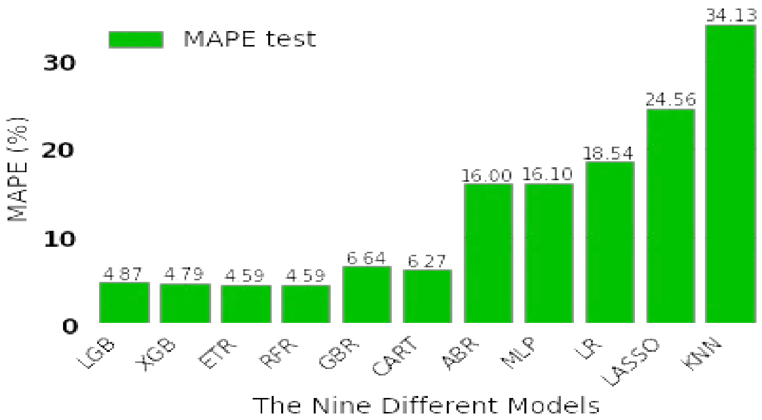

Figure 12. MAPE representation

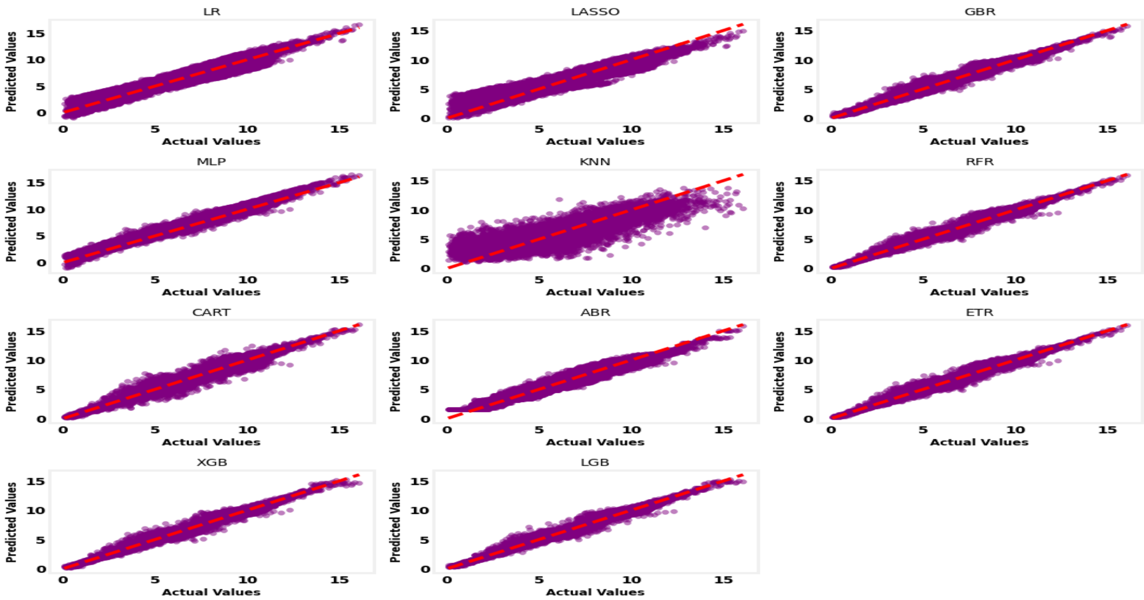

Figure 13. Presentation of point clouds for the different

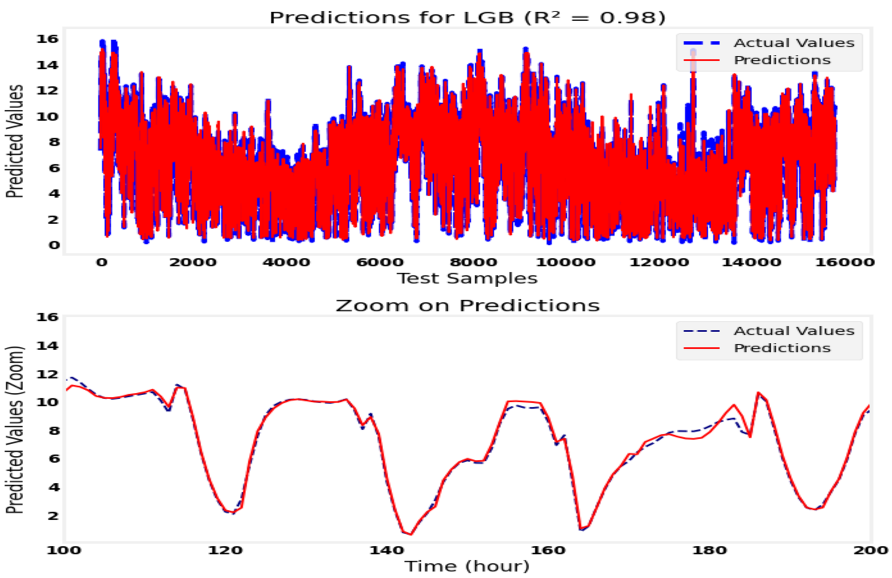

Figure 14. Wind speed prediction profile from the LGB model

10.57647/ijeee.2024.1503.11

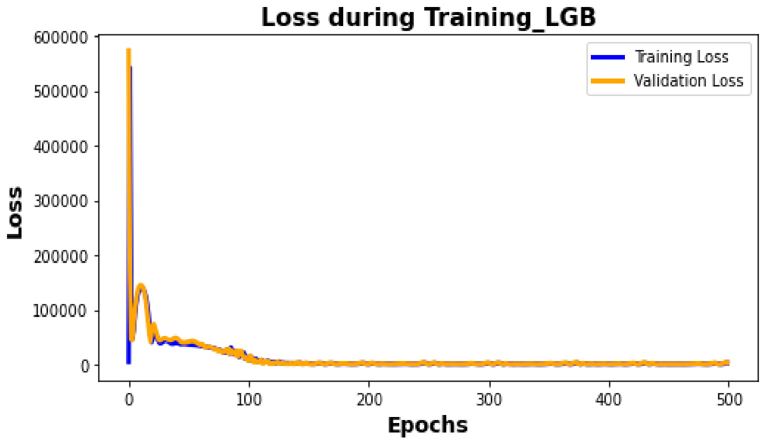

Figure 15. training and validation loss curve

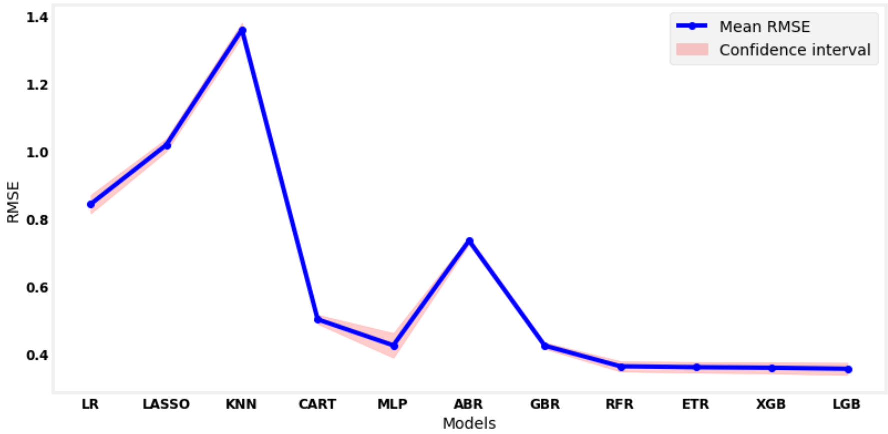

Figure 16. Confidence interval for the different models

Our current study builds on this work by demonstrating the effectiveness of machine learning techniques for predicting wind speed in the city of Amdjarass. In particular, the application of regression models, such as linear regression and the Extra Trees Regressor, reflects the approaches used in other similar research [55, 56].

The performance of these models in capturing wind speed patterns is consistent with their general theoretical understanding. Although similarities can be observed, variations in model performance across different studies highlight the importance of factors such as dataset characteristics, geographical context, and selected machine learning techniques. The superior performance of the Extra Trees Regressor (ETR) model in our study is consistent with the results obtained in similar research [57-59]. This consistency observed across studies suggests that the Extra Trees Regressor (ETR) model is a robust choice for capturing wind speed dynamics in the city of Amdjarass.

3.2.5. Calculation of the power of a wind turbine in Amdjarass

The power calculation is based exclusively on the results of the LGB model. Taking into account the table of adjusted Weibull parameters, the power curve, illustrated in Figure 16, is generated. This curve shows the variation in power over an extended period. A subgraph allows you to zoom in on a shorter period, revealing the power distribution. At a given moment, the wind turbine's power reaches a maximum of 275 kW. In addition, intervals where the power is zero are observed, due to wind speed drops below the wind turbine's start-up speed.

3.2.6. Estimated production for one year

To make the most of the results presented above, the power measured at heights of 10 meters, 55 meters, and 67 meters, as well as the load factor, were calculated. These data are shown in Table 6.

The maximum power values were recorded in February and March, reaching 231.225 kW and 220.732 kW respectively. Conversely, the minimum values were recorded in August and September, with 29.304 kW and 30.393 kW, at a height of 67 meters above ground level. The power measured at heights of 55 meters and 10 meters shows similar characteristics, with peaks in February and March and troughs in August and September. Table 6 shows the load factor (LF) values, with a maximum value of 70% observed in March, indicating optimal power production during that month.

In contrast, the lowest LF values, 14%, were recorded in August and September, which is attributed to low wind speeds during this period.

Figure 17 shows the estimated useful and exploitable power at three specific heights. It appears that the wind turbine produces optimal performance when installed at a height of 67 meters. At 55 meters above ground level, the wind turbine also generates very satisfactory results, reaching power outputs of 190.362 kW in February and 199.411 kW in March, with lower values of 25.272 kW in August and 26.21 kW in September. It is also observed that the actual usable power is significantly lower than the useful power during the months of June, July, August, and September. On the other hand, average useful power is observed in May, October, and November. For the months of January, February, March, April, and December, the power obtained is comparable to the available power.

3.2.7. Analysis of wind turbulence

The average variations observed throughout the year sometimes make it possible to characterize the predominant, or even distinct, wind regimes that regularly influence the site. In Amdjarass, the prevailing winds are mainly the harmattan and the monsoon. Studying the average daily variations makes it possible to determine whether the station will be able to ensure continuity of service, i.e., uninterrupted electricity production.

The analysis of the graph shown in Figure 18 reveals that wind speed is low during certain intervals. In particular, between 600 and 700 hours, there are periods when the wind speed is below average. Therefore, to ensure the continuous use of energy, an alternative source must be used during these fluctuations. The installation of wind turbines in Amdjarass does not guarantee uninterrupted access to energy at all times.

3.3. Wind rose diagrams

Wrplot is a software program for analyzing and visualizing meteorological data. Developed by Lakes Environmental Software, Wrplot allows users to easily create wind rose plots. To determine the location of a wind turbine installation in the study area, WrPlot version 8.0.2 software is used. The wind rose interpretation for the city of Amdjarass is essential for assessing the wind energy potential of this site.

Figure 19 shows a prevailing wind from the east-southeast with a maximum frequency of 40%. The absence of calm winds in this region is a promising indicator for wind energy, as it eliminates the Weibull hybrid distribution often observed in areas where periods of calm winds are frequent. These wind conditions are favorable for the installation of wind turbines. To ensure efficient wind energy production at the site, the wind turbines must be positioned so that their front faces are oriented east-east-south. This allows the blades to effectively capture the prevailing wind.

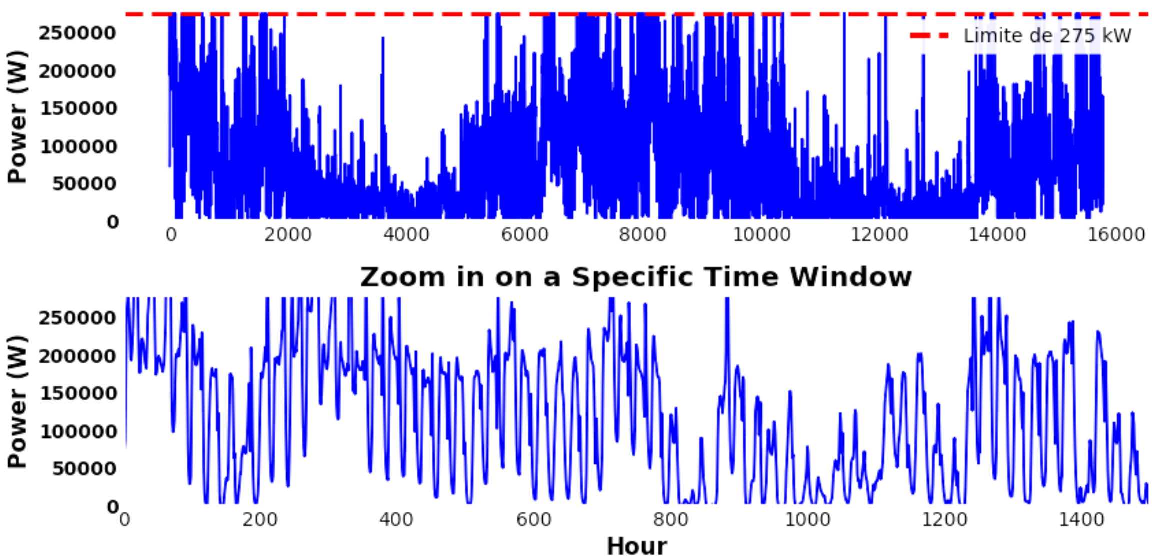

Figure 17. Power curve of the GEV MP 275/32 wind turbine

10.57647/ijeee.2024.1503.11

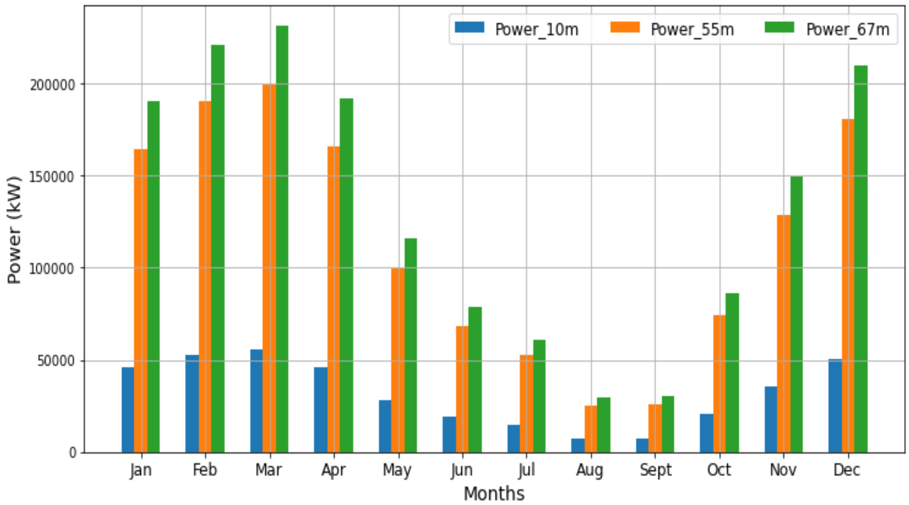

Figure 18. Estimation of useful and exploitable power

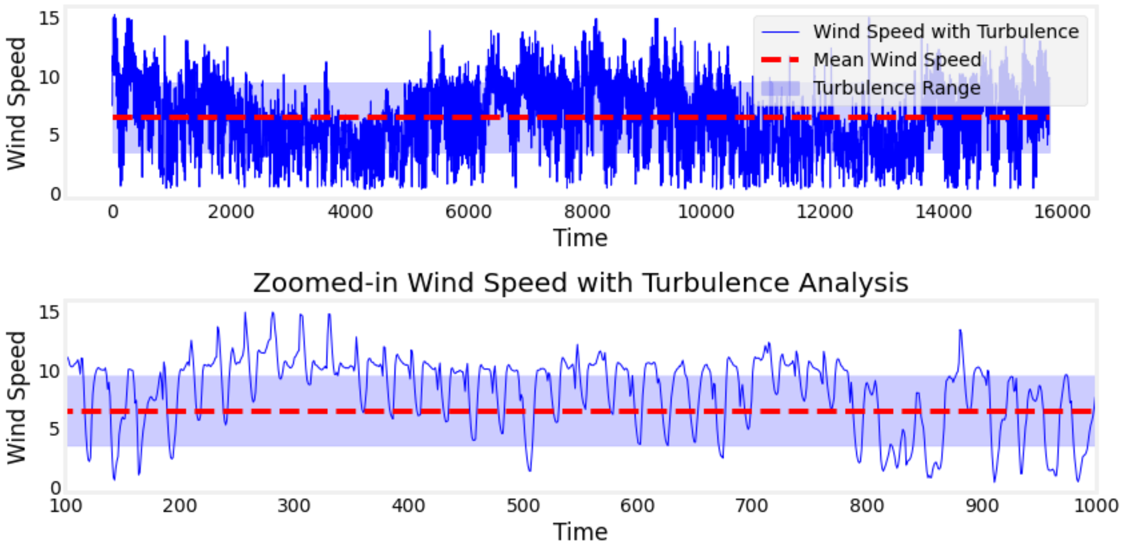

Figure 19. Estimation of useful and exploitable power

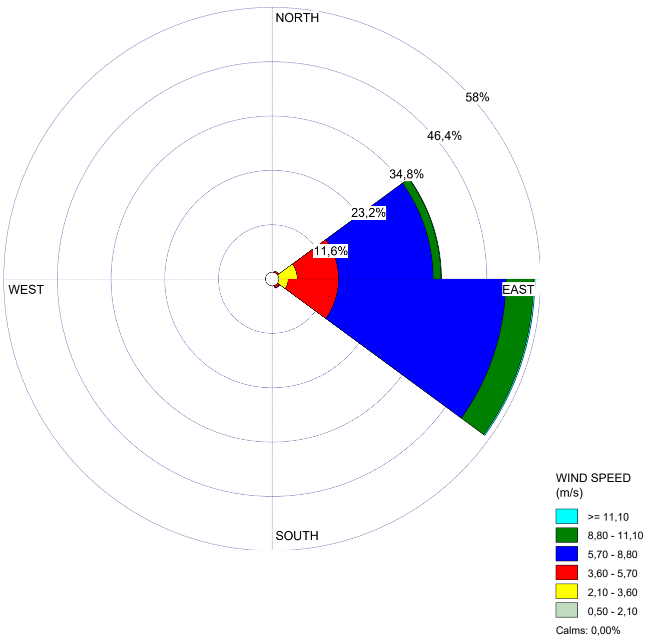

Figure 20. Wind rose diagram

Table 6. Calculation of power at different heights

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
V_W(m/s)	9.81	10.31	10.47	9.85	8.32	7.32	6.71	5.26	5.33	7.55	9.06	10.14
K	3.29	3.37	3.40	3.29	3.03	2.84	2.72	2.41	2.42	2.88	3.16	3.34
C(m/s)	8.80	9.26	9.41	8.83	7.43	6.52	5.97	4.66	4.72	6.73	8.11	9.10
FC	0.64	0.69	0.70	0.65	0.48	0.36	0.29	0.14	0.14	0.39	0.57	0.67
Power_10m (w)	45.679	53.004	55.523	46.130	27.810	18.957	14.592	7.036	7.298	20.761	35.913	50.321
Power_55m (w)	164.057	190.362	199.411	165.67	99.880	68.083	52.410	25.272	26.211	74.563	128.983	180.729
Power_67m (w)	190.231	220.732	231.225	192.108	115.814	78.945	60.771	29.304	30.393	86.459	149.560	209.562

Conclusion

This study evaluates the performance of eleven machine learning algorithms, namely the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm, K-Neighbors Regressor (KNN), Decision Tree Regressor (CART), Multi-Layer Perceptron (MLP), Ada Boost Regressor (ABR), Gradient Boosting Regressor (GBR), Random Forest Regressor (RFR), Extra Trees Regressor (ETR), XGBoost (XGB) and LightGBM (LGB) for predicting wind speed in the city of Amdjarass. Temperature, wind direction, clear sky irradiation, global irradiation, relative humidity, and wind speed were collected at the N'Djamena weather station for a period from January 1, 2015, to December 31, 2023, recorded at 1-hour intervals. To evaluate the effectiveness of these algorithms, three performance metrics were used: R^2, RMSE, and MAE. Among the eleven models, the LGB model performed best with a score of R^2 = 0.986023, RMSE = 0.355507, nRMSE=0.055, MAE = 0.018194 and MAPE=4.8741. This made it possible to calculate the power output. The GEV MP 275/32 wind turbine model was used. It achieves a maximum power output of 231,225 kWh in March and a minimum power output of 29,304 kWh in August. The prevailing wind at the site is from the east-southeast. For good wind energy production, the wind turbines must be positioned facing east-southeast. This allows the blades to capture the prevailing wind efficiently. However, similar studies should be carried out taking into account deep learning for wind prediction.

Funding Statement
The study did not receive funding from any individual or institution.

Authors Contribution
All authors have contributed equally to prepare the paper.

Availability of data and materials
The data used in this study is confidential.

Conflict of interests
The authors declare that there are no conflicts of interest between themselves or with any individual or institution.

References

1. Wang, S., Tang, X., Wang, J., Zhang, B., Sun, W., & Höök, M. (2021). Environmental impacts from conventional and shale gas and oil development in China considering regional differences and well depth. *Resources, Conservation and Recycling*, *167*, 105368.
DOI: https://doi.org/10.1016/j.resconrec.2020.105368

2. Lai, J. P., Chang, Y. M., Chen, C. H., & Pai, P. F. (2020). A survey of machine learning models in renewable energy predictions. Applied Sciences, 10(17), 5975.
DOI: https://doi.org/10.3390/app10175975

3. Panwar, N. L., Kaushik, S. C., & Kothari, S. (2011). Role of renewable energy sources in environmental protection : A review. Renewable and sustainable energy reviews, 15(3), 1513-1524.
DOI: https://doi.org/10.1016/j.rser.2010.11.037

4. Lai, J. P., Chang, Y. M., Chen, C. H., & Pai, P. F. (2020). A survey of machine learning models in renewable energy predictions. Applied Sciences, 10(17), 5975.
DOI: https://doi.org/10.3390/app10175975

5. Tascikaraoglu, A., & Uzunoglu, M. (2014). A review of combined approaches for prediction of short-term wind speed and power. Renewable and Sustainable Energy Reviews, 34, 243-254.
DOI: https://doi.org/10.1016/j.rser.2014.03.033

6. Shang, M., Miao, W., Jia, D. *et al.* Fabrication and Characterization of PMMA/MWCNTs Composite Materials. *J. Wuhan Univ. Technol.-Mat. Sci. Edit.* 38, 1190–1197 (2023).
DOI: https://doi.org/10.1007/s11595-023-2809-y

7. Lazić, L., Pejanović, G., & Živković, M. (2010). Wind forecasts for wind power generation using the Eta model. Renewable Energy, 35(6), 1236-1243.
DOI: https://doi.org/10.1016/j.renene.2009.10.028

8. Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. Renewable energy, 37(1), 1-8.
DOI: https://doi.org/10.1016/j.renene.2011.05.033

9. Lazić, L., Pejanović, G., & Živković, M. (2010). Wind forecasts for wind power generation using the Eta model. Renewable Energy, 35(6), 1236-1243.
DOI: https://doi.org/10.1016/j.renene.2009.10.028

10. Singh, U., Rizwan, M., Alaraj, M., & Alsaidan, I. (2021). A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments. Energies, 14(16), 5196.
DOI: https://doi.org/10.3390/en14165196

11. Wang, X., Guo, P., & Huang, X. (2011). A review of wind power forecasting models. Energy procedia, 12, 770-778.
DOI: https://doi.org/10.1016/j.egypro.2011.10.103

12. Cetinay, H., Kuipers, F. A., & Guven, A. N. (2017). Optimal siting and sizing of wind farms. Renewable Energy, 101, 51-58,
DOI: https://doi.org/10.1016/j.renene.2016.08.008

13. Zhao, X., Bai, M., Yang, X., Liu, J., Yu, D., & Chang, J. (2021). Short-term probabilistic predictions of wind multi-parameter based on one-dimensional convolutional neural network with attention mechanism and multivariate copula distribution estimation. *Energy*, *234*, 121306.
DOI: https://doi.org/10.1016/j.energy.2021.121306

14. Dieudonné, N. T., Armel, T. K. F., Hermann, D. T., Vidal, A. K. C., & René, T. (2023). Optimization of Short-Term Forecast of Electric Power Demand in the city of Yaoundé-Cameroon by a hybrid model based on the combination of neural networks and econometric methods from a designed energy optimization algorithm. Technological Forecasting and Social Change, 187, 122212.
DOI: https://doi.org/10.1016/j.techfore.2022.122212

15. Gu, B., Zhang, T., Meng, H., & Zhang, J. (2021). Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation. Renewable Energy, 164, 687-708.
DOI: https://doi.org/10.1016/j.renene.2020.09.087

16. Dankassoua, M., Madougou, S., Aboubacar, A., & Foulani, A. I. (2017). Study of global solar radiation in Niamey during the pre-monsoon and monsoon periods of 2013 (May to October). Journal of Renewable Energies, 20(1), 131-146.
DOI: https://doi.org/10.54966/jreen.v20i1.615

17. Bounoua, Z., Chahidi, L. O., & Mechaqrane, A. (2021). Estimation of daily global solar radiation using empirical and machine-learning methods: A case study of five Moroccan locations. Sustainable Materials and Technologies, 28, e00261.
DOI: https://doi.org/10.1016/j.susmat.2021.e00261

18. Gu, B., Zhang, T., Meng, H., & Zhang, J. (2021). Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation. Renewable Energy, 164, 687-708.
DOI: https://doi.org/10.1016/j.renene.2020.09.087

19. Gu, B., Zhang, T., Meng, H., & Zhang, J. (2021). Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation. Renewable Energy, 164, 687-708.
DOI: https://doi.org/10.1016/j.renene.2020.09.087

20. Maroufpoor, S., Sanikhani, H., Kisi, O., Deo, R. C., & Yaseen, Z. M. (2019). Long-term modelling of wind speeds using six different heuristic artificial intelligence approaches. International Journal of Climatology, 39(8), 3543-3557.
DOI: https://doi.org/10.1002/joc.6037

21. Salah, S., Alsamamra, H. R., & Shoqeir, J. H. (2022). Exploring wind speed for energy considerations in eastern Jerusalem-Palestine using machine-learning algorithms. Energies, 15(7), 2602.
DOI: https://doi.org/10.3390/en15072602

22. Jain, P., Chhabra, H., Chauhan, U., Singh, D. K., Anwer, T. M. K., Ahammad, S. H., ... & Rashed, A. N. Z. (2023). Multiband metamaterial absorber with absorption prediction by assisted machine learning. *Materials Chemistry and Physics*, *307*, 128180.
DOI: https://doi.org/10.1016/j.matchemphys.2023.128180

23. Hanoon, M. S., Ahmed, A. N., Kumar, P., Razzaq, A., Zaini, N. A., Huang, Y. F., ... & El-Shafie, A. (2022). Wind speed prediction over Malaysia using various machine learning models: potential renewable energy source. *Engineering Applications of Computational Fluid Mechanics*, *16*(1), 1673-1689.
DOI: https://doi.org/10.1080/19942060.2022.2103588

24. Suthar, G., Singhal, R. P., Khandelwal, S., & Kaul, N. (2023). Spatiotemporal variation of air pollutants and their relationship with land surface temperature in Bengaluru, India. *Remote Sensing Applications: Society and Environment*, *32*, 101011.
DOI: https://doi.org/10.1016/j.rsase.2023.101011

25. Khosravi, A. K. R. M. L. P. J., Koury, R. N. N., Machado, L., & Pabon, J. J. G. (2018). Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system. Sustainable Energy Technologies and Assessments, 25, 146-160.
DOI: https://doi.org/10.1016/j.seta.2018.01.001

26. Demolli, H., Dokuz, A. S., Ecemis, A., & Gokcek, M. (2019). Wind power forecasting based on daily wind speed data using machine learning algorithms. Energy Conversion and Management, 198, 111823.
DOI: https://doi.org/10.1016/j.enconman.2019.111823

27. Routray, A., Mistry, K. D., Arya, S. R., & Chittibabu, B. (2021). Applied machine learning in wind speed prediction and loss minimization in unbalanced radial distribution system. Energy sources, Part A: recovery, utilization, and environmental effects, 1-21.
DOI: https://doi.org/10.1080/15567036.2020.1859010

28. Irfan, A. S. M., Bhuiyan, N. H., Hasan, M., & Khan, M. M. (2021, July). Performance analysis of machine learning techniques for wind speed prediction. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.
DOI: https://doi.org/10.1109/ICCCNT51525.2021.9579564

29. Mogos, A. S., Salauddin, M., Liang, X., & Chung, C. Y. (2021, September). Very short-term wind speed prediction techniques using machine learning. In 2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) (pp. 1-6). IEEE.
DOI: https://doi.org/10.1109/CCECE53047.2021.9569134

30. Wang, T. (2021). A combined model for short-term wind speed forecasting based on empirical mode decomposition, feature selection, support vector regression and cross-validated lasso. PeerJ Computer Science, 7, e732.
DOI: https://doi.org/10.7717/peerj-cs.732

31. Nediguina, M. K., Abdraman, M. A., Barka, M., & Tahir, A. M. (2022). Electric Water Pumping Powered by a Wind Turbine in North East Chad. World Journal of Applied Physics, 7(2), 21-31.
DOI: https://doi.org/10.11648/j.wjap.20220702.12

32. Guher, AB, Tasdemir, S., & Yaniktepe, B. (2020). Efficient estimation of hourly global solar radiation using machine learning algorithms. International Journal of Solar Energy. 2020 (1), 8843620.
DOI: https://doi.org/10.1155/2020/8843620

33. Sarker, IH Machine Learning : algorithmes, applications concrètes et orientations de recherche. Sn comput. Sci. 2 , 160 (2021).
DOI: https://doi.org/10.1007/s42979-021-00592-x

34. Herm, L. V., Heinrich, K., Wanner, J., & Janiesch, C. (2023). Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. International Journal of Information Management, 69, 102538.
DOI: https://doi.org/10.1016/j.ijinfomgt.2022.102538

35. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj computer science, 7, e623.
DOI: https://doi.org/10.7717/peerj-cs.623

36. Chicco, D., Warrens, M. J., & Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation.
DOI: http://dx.doi.org/10.7717/peerj-cs.623

37. Kvalheim, O. M., Arneberg, R., Grung, B., & Rajalahti, T. (2018). Determination of optimum number of components in partial least squares regression from distributions of the root-mean-squared error obtained by Monte Carlo resampling. Journal of Chemometrics, 32(4), e2993.
DOI: https://doi.org/10.1002/cem.2993

38. Pati, N., Gourisaria, M. K., Das, H., & Banik, D. (2023, April). Wind speed prediction using machine learning techniques. In *2023 11th International conference on emerging trends in engineering & technology-signal and information processing (ICETET-SIP)* (pp. 1-6). IEEE.
DOI: https://doi.org/10.1109/ICETET-SIP58143.2023.10151597

39. Demolli, H., Dokuz, A. S., Ecemis, A., & Gokcek, M. (2019). Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Conversion and Management*, *198*, 111823.
DOI: https://doi.org/10.1016/j.enconman.2019.111823

40. Manoj, S. O., Ananth, J. P., Rohini, M., Dhanka, B., Pooranam, N., & Arumugam, S. R. (2022). FWS-DL: forecasting wind speed based on deep learning algorithms. In *Artificial intelligence for renewable energy systems* (pp. 353-374). Woodhead Publishing.
DOI: https://doi.org/10.1016/B978-0-323-90396-7.00007-9

41. Ho, TK (1995, août). Forêts de décision aléatoires. Dans les actes de la 3e conférence internationale sur l'analyse et la reconnaissance de documents (vol. 1, pp. 278-282). IEEE.
DOI: https://doi.org/10.1109/ICDAR.1995.598994

42. Thakor, S. G., Rana, V. A., Vankar, H. P., & Pandit, T. R. (2021). Dielectric spectroscopy and structural characterization of nano-filler-loaded epoxy resin. *Journal of Advanced Dielectrics*, *11*(02), 2150011.
DOI: https://doi.org/10.1142/S2010135X21500119

43. Vovchenko, L. L., Lozitsky, O. V., Matzui, L. Y., Oliynyk, V. V., Zagorodnii, V. V., Milovanov, Y. S., ... & Prylutskyy, Y. I. (2020). Impedance characterization and microwave permittivity of multi-walled carbon nanotubes/BaTiO3/epoxy composites. *Applied Physics A*, *126*(10), 801.
DOI: https://doi.org/10.1007/s00339-020-03831-9

44. Kvalheim, O. M., Arneberg, R., Grung, B., & Rajalahti, T. (2018). Determination of optimum number of components in partial least squares regression from distributions of the root-mean-squared error obtained by Monte Carlo resampling. Journal of Chemometrics, 32(4), e2993.
DOI: https://doi.org/10.1002/cem.2993

45. Sajid, S. W., Hasan, M., Rabbi, M. F., & Abedin, M. Z. (2023). An ensemble LGBM (light gradient boosting machine) approach for crude oil price prediction. In *Novel financial applications of machine learning and deep learning: algorithms, product modeling, and applications* (pp. 153-165). Cham: Springer International Publishing.
DOI: https://doi.org/10.1007/978-3-031-18552-6_9

46. Fan, J., Wang, X., Wu, L., Zhang, F., Bai, H., Lu, X., & Xiang, Y. (2018). New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: a case study in South China. Energy conversion and management, 156, 618-625.
DOI: https://doi.org/10.1016/j.enconman.2017.11.085

47. Mohsen-Nia, M., Amiri, H., & Jazi, B. (2010). Dielectric constants of water, methanol, ethanol, butanol and acetone: measurement and computational study. *Journal of Solution Chemistry*, *39*(5), 701-708.
DOI: https://doi.org/10.1007/s10953-010-9538-5

48. Rehman, S. (1998). Solar radiation over Saudi Arabia and comparisons with empirical models. Energy, 23(12), 1077-1082.
DOI: https://doi.org/10.1016/S0360-5442(98)00057-7

49. Nematchoua, M. K., Orosa, J. A., & Afaifia, M. (2022). Prediction of daily global solar radiation and air temperature using six machine learning algorithms; a case of 27 European countries. Ecological Informatics, 69, 101643.
DOI: https://doi.org/10.1016/j.ecoinf.2022.101643

50. Fan, J., Wang, X., Wu, L., Zhang, F., Bai, H., Lu, X., & Xiang, Y. (2018). New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: a case study in South China. Energy conversion and management, 156, 618-625.
DOI: https://doi.org/10.1016/j.enconman.2017.11.085

51. Harris, R. I., & Cook, N. J. (2014). The parent wind speed distribution: Why Weibull?. Journal of wind engineering and industrial aerodynamics, 131, 72-87.
DOI: https://doi.org/10.1016/j.jweia.2014.05.005

52. Satwika, N. A., Hantoro, R., Septyaningrum, E., & Mahmashani, A. W. (2019). Analysis of wind energy potential and wind energy development to evaluate performance of wind turbine installation in Bali, Indonesia. Journal of Mechanical Engineering and Sciences, 13(1), 4461-4476.
DOI: https://doi.org/10.15282/jmes.13.1.2019.09.0379

53. Boro, D., Donnou, H. E. V., Kossi, I., Bado, N., Kieno, F. P., & Bathiebo, J. (2019). Vertical profile of wind speed in the atmospheric boundary layer and assessment of wind resource on the Bobo Dioulasso site in Burkina Faso.
DOI: https://doi.org/10.4236/sgre.2019.1011016

54. Soulouknga, M. H., Oyedepo, S. O., Doka, S. Y., & Kofane, T. C. (2020). Evaluation of the cost of producing wind-generated electricity in Chad. International Journal of Energy and Environmental Engineering,11, 275-287.
DOI: https://doi.org/10.1007/s40095-019-00335-y

55. Adjiri, S., Nedjari, H. D., & Haddouche, S. (2013). Estimation du potentiel éolien pour la région d'In Salah. Journal of Renewable Energies, 16(4), 657-666.
DOI: https://doi.org/10.54966/jreen.v16i4.408

56. Becker, R., & Thrän, D. (2018). Optimal siting of wind farms in wind energy dominated power systems. Energies, 11(4), 978.
DOI: https://doi.org/10.3390/en11040978

57. Islam, M. R., Rahman, M. N., & Mannan, M. A. (2016). Study of wind power in different parts of Bangladesh. http://dspace.aiub.edu:8080/jspui/handle/123456789/2062

58. Alves, D., Mendonça, F., Mostafa, S. S., & Morgado-Dias, F. (2023). The potential of machine learning for wind speed and direction short-term forecasting: a systematic review. Computers, 12(10), 206.
DOI: https://doi.org/10.3390/computers12100206

59. Fu, L., Li, J., & Chen, Y. (2023). An innovative decision making method for air quality monitoring based on big data-assisted artificial intelligence technique. *Journal of Innovation & Knowledge*, *8*(2), 100294.
DOI: https://doi.org/10.1016/j.jik.2022.100294