Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/130414

Comparative analysis of the predictive capabilities of some machine learning models: A case study using wind speed data

Kgothatso Makubyane[1,2] *and Caston Sigauke*[1,*]

[1]Department of Mathematical and Computational Sciences, University of Venda, Private Bag X5050, Thohoyandou 0950, Limpopo, South Africa
[2]DST-CSIR National e-Science Postgraduate Teaching and Training Platform (NEPTTP), Mathematical Sciences Building, West Campus, Private Bag 3, Wits 2050, Gauteng, South Africa

Received: 17 Aug. 2023, Revised: 3 Apr. 2024, Accepted: 10 Apr. 2024
Published online: 1 Jul. 2024

Abstract: The widespread use of fossil fuels for global energy production significantly contributes to global warming. This study presents a comparative analysis of various machine learning models, which are the long short-term memory (LSTM) network, support vector regression (SVR), and gradient boosting method (GBM). Gaussian process regression (GPR) is a benchmark model across different forecasting horizons. The study uses South African wind speed data from 1 January 2018 to 31 December 2021, sourced from the Western Cape province. The dataset underwent preprocessing, and diverse feature selection techniques were implemented to enhance model accuracy. Performance evaluation of the models was done using mean absolute error (MAE), root mean squared error (RMSE), and mean absolute scaled error (MASE). Results indicate that SVR exhibits superior accuracy to other models for two distinct forecast horizons ($h = 670$ and $h = 1339$), respectively. Additionally, GPR surpasses other models for the forecasting horizon $h = 224$. This study provides insights into the comparative strengths and weaknesses of different machine learning models for wind speed prediction, which could be useful in selecting an appropriate model for future applications in renewable energy and weather forecasting. Potential areas for future research include improving prediction accuracy via ensemble deep learning algorithms and incorporating additional meteorological variables. Moreover, investigating temporal dynamics, broadening geographical coverage and integrating uncertainty quantification methods can improve wind speed prediction, thereby facilitating more effective renewable energy planning and decision-making processes

Keywords: LSTM; Predictive capability; Renewable energy; SVR; Skill score.

1 Introduction

Accurate wind speed prediction is essential for numerous purposes, including wind energy generation, weather prediction, and monitoring air quality. Machine learning models have shown great potential for predicting wind speed due to their ability to learn complex patterns and relationships in data. However, with the increasing number of machine learning algorithms available, it is challenging to determine which model performs the best for a particular task.

Therefore, in this study, we aim to compare the predictive capabilities of several machine learning models using wind speed data. Specifically, we will evaluate the performance of multiple models, including Long Short-Term Memory (LSTM), Support Vector Regression (SVR), Stochastic Gradient Boosting Method (SGBM) and Gaussian Process Regression (GPR). By comparing the performance of these machine learning models, we aim to identify the most accurate and reliable model for predicting wind speed on different forecasting horizons. This information can improve wind energy production efficiency and forecasting accuracy.

* Corresponding author e-mail: caston.sigauke@univen.ac.za

1.1 Background

According to the Council for Scientific and Industrial Research (CSIR), South Africa (SA) is experiencing the worst year of load-shedding. However, there is a solution to this obstacle: Renewable energy resources (Wind, Sun and Water). The ongoing dependence on fossil fuels (such as coal, oil, and natural gas) for energy generation worsens environmental problems such as global warming and associated issues. Burning fossil fuels in energy production has led to adverse environmental effects, such as adverse climate patterns and impacts on human health (Sun et al. [1]). Patel [2] examined the environmental and economic implications of fossil fuels. These challenges have spurred numerous researchers to explore wind energy as a potential alternative electricity source. Wind energy generation relies heavily on meteorological factors such as wind speed, atmospheric pressure, humidity, and temperature. However, due to the uncontrollable nature of wind, regulating the amount of electrical energy produced from wind is exceedingly difficult (Patel [2]).

1.2 A survey of related literature

In recent years, there has been a rise in the use of machine learning models for wind speed prediction owing to their capacity to manage complex data patterns and relationships. Numerous investigations have been conducted to assess the efficacy of various machine learning models. One notable study conducted by Mishra et al. [3] scrutinised the performance of five distinct models: Deep Feed Forward (DFF), Deep Convolutional Network (DCN), Recurrent Neural Network (RNN), Attention mechanism (Attention), and Long Short-Term Memory Networks (LSTM) in predicting wind speed data. The study's findings indicated that the Attention and DCN models outperformed others when applied with Wavelet or FFT signal preprocessing, while some models demonstrated superior performance without any preprocessing. Another study by Elsaraiti and Merabet [4] compared the predictive capabilities of ARIMA and LSTM models for wind speed prediction. Their investigation revealed that the LSTM method exhibited greater accuracy than ARIMA, as assessed by the root mean square error (RMSE) metric.

In a separate study, Dhakal et al. [5] conducted a comparative analysis of various models including Weibull probability density-based WSP (WEB), Rayleigh probability density-based WSP (RYM), autoregressive integrated moving average (ARIMA), Kalman filter, support vector machines (SVR), and artificial neural network (ANN), alongside hybrid models, for short-term wind speed forecasting. The researchers introduced an error correction algorithm for the probability density-based wind speed prediction model to enhance prediction accuracy. Their findings indicate a notable enhancement in the performance of wind speed prediction models.

Similarly, Mutavhatsindi et al.[6] investigated the predictive capabilities of Feed Forward Neural Networks (FFNN), LSTM networks, and SVR models for short-term solar energy forecasting. Accurate wind speed forecasting is crucial for effectively implementing wind power generation to uphold power system stability. This paper aims to raise awareness among government officials regarding the advantages of renewable energy production, aiding governmental authorities and decision-makers in effectively managing and mitigating the impacts of global warming. In South Africa, Odhiambo [7] found a correlation between electricity usage and economic expansion that is directly proportional.

Alkesaiberi et al. [8] conducted a comparative examination of Gaussian Process Regression (GPR), Support Vector Regression (SVR) employing various kernels, and ensemble learning (ES) models, specifically Boosted trees and Bagged trees, for wind power forecasting. Wind power data from three different locations were utilised to evaluate the efficacy of these models. Empirical findings from this investigation demonstrated that both ensemble and GPR models outperformed the other methods. Buturache and Stancu [9] carried out a comprehensive comparative study on predicting wind energy, examining Artificial Neural Networks, Support Vector Regression, Random Trees, and Random Forest. The authors also discussed the advantages and disadvantages of the proposed models. In another study, Singh and Rizwan [10] used machine learning models for short-term wind power prediction, using Support Vector Regression (SVR) and Gradient Boosting Regression Trees (GBRT). GBRT exhibited superior performance over the SVR model based on various evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

Tarek et al. [11] developed an optimisation technique named Stochastic Fractal Search and Particle Swarm Optimization (SFSPSO) to optimise parameters of the Long Short-Term Memory (LSTM) network for short-term wind power forecasting. They then compared the predictive capabilities of the Deep Neural Network (DNN), K-Nearest Neighbor (KNN) regressor, LSTM, Averaging model, Random Forest (RF) regressor, Bagging regressor, and Gradient Boosting (GB) regressor with the proposed SFSPSO method. The SFSPSO method outperformed the base models across five predictive evaluation metrics. Daniel et al. [12] conducted a related study by comparing short-term prediction of average wind speed data using statistical and machine learning models. They combined predictions from individual models using an additive quantile regression averaging method, resulting in significantly improved forecast accuracy.

In summary, the results of these studies suggest that the choice of a machine learning model for predicting wind speed depends on the specific dataset and the performance metrics of interest. However, none of the studies considered different forecasting horizons to assess the models' predictive abilities. It is this shortcoming that the current study seeks to correct by considering the predictive skills and capabilities of some machine learning models for forecasting the maximum wind speed from one location at different forecasting horizons. These models can improve wind energy production efficiency and forecasting accuracy.

1.3 Research highlights

Based on the insights from the literature review outlined in subsection 1.2, this study offers valuable contributions to local and global wind speed forecasting for power generation. The Augmented Dickey-Fuller (ADF) test revealed that the focused period data is constant over time, indicating that its stationary nature is crucial for reliable forecasting. The Lasso investigation highlighted the significant impact of atmospheric pressure and air temperature on accurate wind speed prediction. This finding is particularly relevant as wind turbines operate optimally under warm air conditions and low barometric pressure, facilitating faster rotation. Moreover, across various forecasting models, including LSTM, SVR, SGBM, and SGB, our study identified SVR as the most effective model across forecasting horizons h=224, 670, and 1339. These findings highlight the necessity of factoring in atmospheric conditions when developing wind speed prediction models to improve their precision and suitability for renewable energy systems.

The remainder of the paper is structured as follows: Section 2 outlines the methodologies employed in the study, which include LSTM networks and support vector regression. Additionally, this section provides an overview of feature selection techniques and methods for evaluating model predictive performance. Empirical findings are presented in Section 3, with concluding remarks presented in Section 4.

2 Methods

The section summarises the models, including a benchmark model used in the study.

2.1 Support vector regression

Support vector regression (SVR), an extension of the support vector machines (SVM) model introduced by Drucker et al. [13], works on the premise of finding a function $f(r_i)$ that predicts the output variable q from the input variable r, but in a higher-dimensional feature space. Given training data with input values $(r_1, r_2, ..., r_t)$ and output values $(q_1, q_2, ..., q_t)$, the SVR algorithm aims to derive this function. SVM models leverage various fundamental kernel functions, such as sigmoid, linear, polynomial, normal distribution, radial basis function, and quadratic radial basis function (Zendehboudi [14]). The formulation of the Support Vector Regression function is presented in equation (1), which is

$$f(x) = \omega\phi(r) + b. \tag{1}$$

From equation (1), ω represents the weight vector, b represents the bias term, and $\phi(r)$ stands for a predetermined mapping function for inputs r.

In order to determine the values of ω and b, it is crucial to minimise the regularised risk function $R(f)$, which can be formulated as follows:

$$R(f) = \frac{1}{2}||\omega||^2 + J\frac{1}{t}\sum_{i=1}^{t} L_\varepsilon(q_i, f(r_i)), \tag{2}$$

In equation (2), $||\omega||^2$ represents a regularisation term designed to maintain the function's capability constant, and J denotes the error in the cost function. Expressing the second term from equation (2), we have:

$$L_\varepsilon(q_i, f(r_i)) = \{|q_i = f(r_i)| = \varepsilon, |q_i = f(r_i)| \geq \varepsilon\}. \tag{3}$$

Equation (2) introduced a method for determining the parameters ω and b by presenting a favourable loose attribute $\xi_i(^*)$, thereby demonstrating the transformation of the fundamental objective function outlined in equation (4).

$$\min \frac{1}{2}||\omega||^2 + J\frac{1}{l}\sum_{i=1}^{l}(\varepsilon_i + \xi_i^*)) \tag{4}$$

$$\alpha(r) = \begin{cases} q_i - < \omega, r_i > - \geq \varepsilon + \xi_i \\ < \omega, r_i > + b - q_i \geq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{5}$$

Using Lagrange multipliers can lead to a quicker solution for Equation (2).

2.2 Long short-term memory

The LSTM network, proposed by Hochreiter and Schmidhuber [15], represents a significant advancement in Recurrent Neural Network (RNN) architecture. Unlike traditional RNNs, which suffer from the loss of past information, leading to challenges in learning long-distance dependencies, LSTM addresses this limitation by facilitating the retention of information over extended periods. In practical terms, LSTM retains information over prolonged durations, making it particularly suitable for tasks requiring long-term memory processing. Moreover, in the context of time series forecasting, LSTM introduces several gates that enhance memory retention, as demonstrated by Hochreiter and Schmidhuber [15].

Let $(x_1, x_2, \ldots, x_t)$ represent the input values and $(y_1, y_2, \ldots, y_t)$ the corresponding output values of historical data to be forecasted. The following foundational system equations can characterise the LSTM network:

$$f_t = g(W_t.[x_t, h_{t-1} + b_f]), \tag{6}$$

$$i_t = g(W_i.[x_t, h_{t-1} + b_i]), \tag{7}$$

$$O_t = g(W_o.[x_t, h_{t-1} + b_o]), \tag{8}$$

$$C_t = f_t c_{t-1} + i_t \tanh(g(W_o.[x_t, h_{t-1} + b_c])), \tag{9}$$

$$h_t = O_t \tanh(c_t). \tag{10}$$

Figure 1 shows the basic architecture of the LSTM network.

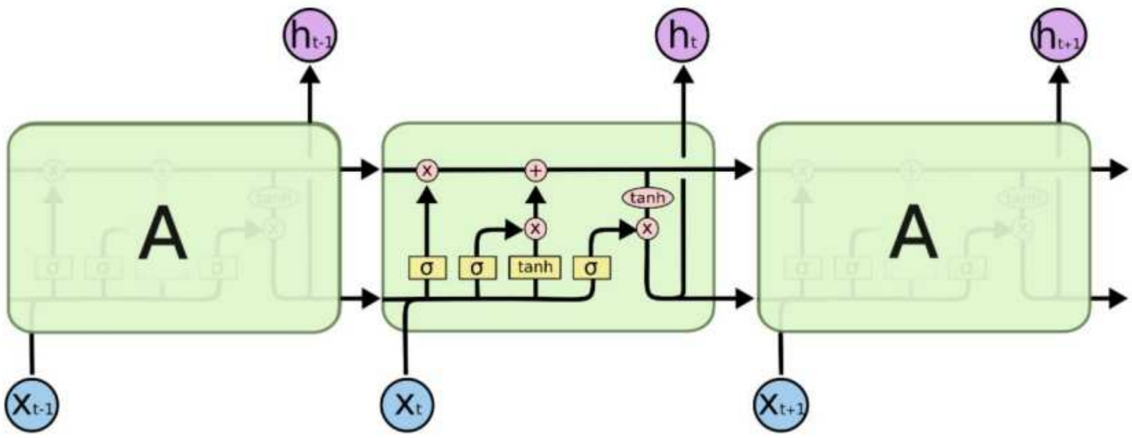

Fig. 1: Basic architecture of the LSTM network Yu et al. [16].

2.3 Stochastic gradient boosting

Stochastic gradient boosting (SGB) is a powerful machine learning method that has become increasingly popular in recent years, particularly in the context of predictive modelling and data analysis (Friedman [17]). This method is a variant of gradient boosting, a popular ensemble method combining multiple weak models to create a more accurate predictive model. The main difference between gradient boosting and stochastic gradient boosting is that the latter introduces randomness into the algorithm by randomly sampling subsets of the training data for each iteration. This approach can help reduce overfitting and improve the model's generalizability, particularly when dealing with large datasets (Friedman [17]). The SGB model is given in equation (11) as,

$$F(x) = \sum_{m=1}^{M} \beta_m h(x;\gamma_m), \tag{11}$$

where $h(x;\gamma_m) \in \mathbb{R}$ are functions of x with parameters γ_m and β_m which limit over fitting (Friedman [17], Hastie et al. [18]).

2.4 Benchmark model: Gaussian process regression

Gaussian process regression (GPR) is an important machine learning technique used extensively across industrial and academic domains. It is a non-parametric algorithm that can model complex nonlinear relationships between variables and provides probabilistic predictions. While it can be computationally expensive for large datasets, several techniques are available to reduce its computational complexity. The key advantage of Gaussian process regression is its ability to model complex nonlinear relationships between variables without assuming any specific parametric form of the underlying distribution. This allows the algorithm to handle many data types, including continuous, discrete, and categorical. A GPR is given in equation (12) as,

$$f(X) \sim GP(m(x), k(x,x^1)). \tag{12}$$

In (12), X represents a collection of independent variables $x_1, x_2, ..., x_n$. The function $f(X)$ corresponds to applying the function f individually to each variable in X, resulting in $f(x_1), f(x_2), ..., f(x_n)$. Two data points within the variable set X are denoted as x and x^1. The function $f(X)$ represents the dependent variable, and $m(x)$ denotes the mean function, where $E(f(x))$ equals $m(x)$. The covariance matrix, $k(x,x^1)$, referred to as the kernel function, is employed. Various kernel functions can be utilised in Gaussian Process Regression (GPR), but in this study, we focus on employing the radial basis kernel function.

2.5 Feature selection

This section discusses feature selection using Lasso to select only significant features for forecasting wind speed. This study uses the Lasso feature selection approach, introduced by Bien et al. [19]. Several variable selection strategies exist, but this study will focus on Lasso. Lasso reduces model complexity by selecting the most significant variables for predicting the output feature. This is crucial for preventing overfitting, enhancing model clarity, and reducing computational time. The Lasso formula is represented by equation (13), which is as follows:

$$\text{Lasso} = \min_{\beta} \frac{1}{n}(Y - X\beta) + \Gamma \sum_{i=1}^{m} ||\beta_i||. \tag{13}$$

In equation (13), n denotes the length of the data points, and $\Gamma \geq 0$ signifies the penalty strength components. As introduced by Tibshirani (1996), the Lasso technique utilises a subset of predictors, leading to more straightforward and easily understandable models.

2.6 Evaluation metrics

To evaluate the accuracy of our predictive models, we will employ metrics including the mean absolute error (MAE), relative MAE (rMAE), root mean square error (RMSE), relative RMSE (rRMSE), mean bias error (MBE), and mean absolute scaled error (MASE)." The mathematical representation for the above metrics are respectively given as,

$$\text{MAE} = \frac{1}{m} \sum_{t=1}^{m} |y_t - \hat{y}_t|, \tag{14}$$

$$\text{rMAE} = \frac{1}{m}\sum_{t=1}^{m}[\frac{y_t - \hat{y}_t}{y_t}], \tag{15}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}(y_t - \hat{y}_t)^2}{m}}, \tag{16}$$

$$\text{rRMSE} = \frac{100}{\bar{y}}\sqrt{\frac{\sum_{t=1}^{m}(y_t - \hat{y}_t)^2}{m}}, \tag{17}$$

$$\text{MBE} = \frac{1}{m}\sum_{t=1}^{m}(y_t - \hat{y}_t), \tag{18}$$

$$\text{MASE} = \frac{y_t - \hat{y}_t}{\frac{1}{m-1}\sum_{i=2}^{m}|y_i - y_{i-1}|}, \tag{19}$$

where y_t represents the actual values at time t, $\hat{y}_t$ denotes the predicted model values at time t, $\bar{y}$ signifies the mean of all actual observations, and n stands for the sample size of the total observations. The Mean Bias Error (MBE) shares the same units as the dependent variable. Positive MBE values indicate underestimation, while negative values indicate overestimation.

The mean absolute scaled error (MASE) is particularly suitable for analysing time series data, offering insight into the predictive effectiveness relative to simple naive predictions. A MASE value of 1 signifies that the proposed model performs as well as the naive model. MASE values less than 1 indicate superior performance compared to the naive model, while values greater than 1 suggest inferior performance.

2.7 Predictive skill and capability

2.7.1 Predictive skill

In this study, we use MASE as outlined in equation (19) to assess the predictive capability. ($\text{PS}_j, j = 1, \ldots k$) where k is the number of proposed models. This helps benchmark the quality of the model's predictions to the benchmark model (the GPR model in this study). The prediction skill is given in equation (20) as,

$$\text{PS}_j = \left(1 - \frac{\text{MASE}_j}{\text{MASE}_{\text{benchmark}}}\right) \times 100, \tag{20}$$

where PS_j is a measure of the superiority of the proposed model's prediction to the benchmark model's prediction. The model giving the greatest positive percentage change is considered the superior model.

2.7.2 Assessing the predictive capability using the Giacomini-White test

The Giacomini-White (GW) test generalises the Diebold-Mariano (DM) test (Giacomini [21]). It is a test on equal conditional predictive ability. The DM and GW tests are applicable to both nested and non-nested models. Nevertheless, the GW test offers an additional benefit by incorporating uncertainty in parameter estimation (Lago [22]). It evaluates the conditional predictive ability of competing forecast pairs. The test is based on the regression presented in equation (21), which can be expressed as follows:

$$\Delta_d^{f,h} = \phi^1 \mathbb{X}_{d-1} + \varepsilon_d, \tag{21}$$

where $\mathbb{X}_{d-1}$ has elements from the information set on day $d-1$, with the pair of forecasts from models f and h. For a detailed discussion of the GW test, see Giacomini [21] and Lago [22], among others.

3 Empirical results

This section presents the empirical results and discusses the results obtained from the study. Python and R programming software were used to obtain the empirical results. The primary libraries used for analysis in R include ggplot2, tseries, and magic. Ihaka and Gentleman [23] are credited as the creators of the R programming language, which is presently under development by Team [24]. In Python, the principal packages utilised for analysis include NumPy, pandas, matplotlib, Seaborn, Keras, and TensorFlow. Van Rossum developed this language [25].

3.1 Data and study area

The study area WM07 (Beaufort West) with coordinates 32°58′0.2″S and 22°33′23.8″E is shown in Figure 2.

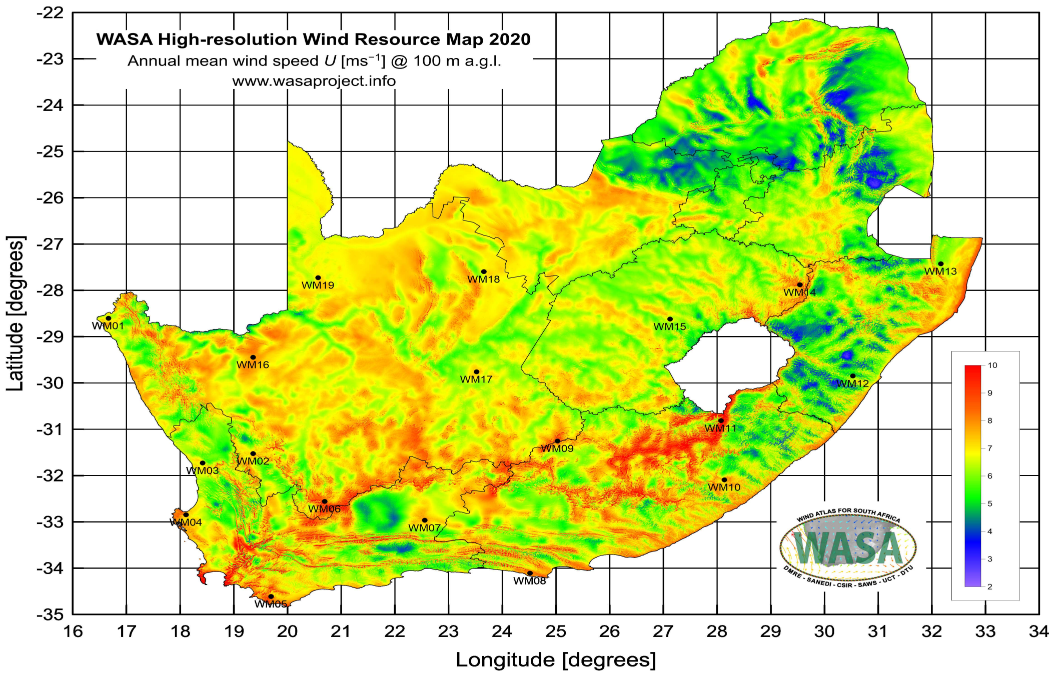

Fig. 2: WASA High-Resolution Wind Resource Map. Source: https://www.wasaproject.info/.

3.2 Exploratory data analysis

Testing for stationarity using the Augmented Dickey-Fuller test

The Augmented Dickey-Fuller (ADF) test is used to assess the stationarity of wind speed recorded at location WM07 (Beaufort West) in the Western Cape Province, specifically at 62 meters above sea level. Table 1 presents the summary statistics for this wind speed data. From 1 December 2021 to 1 January 2022, the wind speed ranged from a minimum of 1.129 to a maximum of 26.598 meters per second at this altitude. The mean wind speed was calculated to be 8.951, while the median was found to be 8.581, providing insight into the central tendency of the data. Additionally, the skewness value of 0.331 and the kurtosis value of 2.993 indicate that the wind speed distribution at 62 meters above sea level is positively skewed and exhibits mesokurtosis, suggesting a departure from normality.

Table 1: Descriptive Statistics.

	min	$1^{st}Qu$	median	mean	$3^{st}Qu$	max	skewness	kurtosis
Wind_speed	1.129	6.718	8.581	8.951	11.065	26.589	0.331	2.993

Fig. 3 illustrates the wind speed recorded in the Western Cape Province at 62m above sea level from 1 December 2021 to 1 January 2022. There appears to be no discernible upward or downward trend in the data during this time frame. However, an observable pattern of approximately a 15% increase occurs approximately every four days, indicating a seasonal influence. Notably, there is a notable peak in wind speed between December 25$^{\text{th}}$, 2021, and December 29$^{\text{th}}$, 2021, possibly attributable to particularly windy conditions during that period. Furthermore, the graph in Figure 3 illustrates that the average wind speed remains relatively constant at approximately 8m/s.

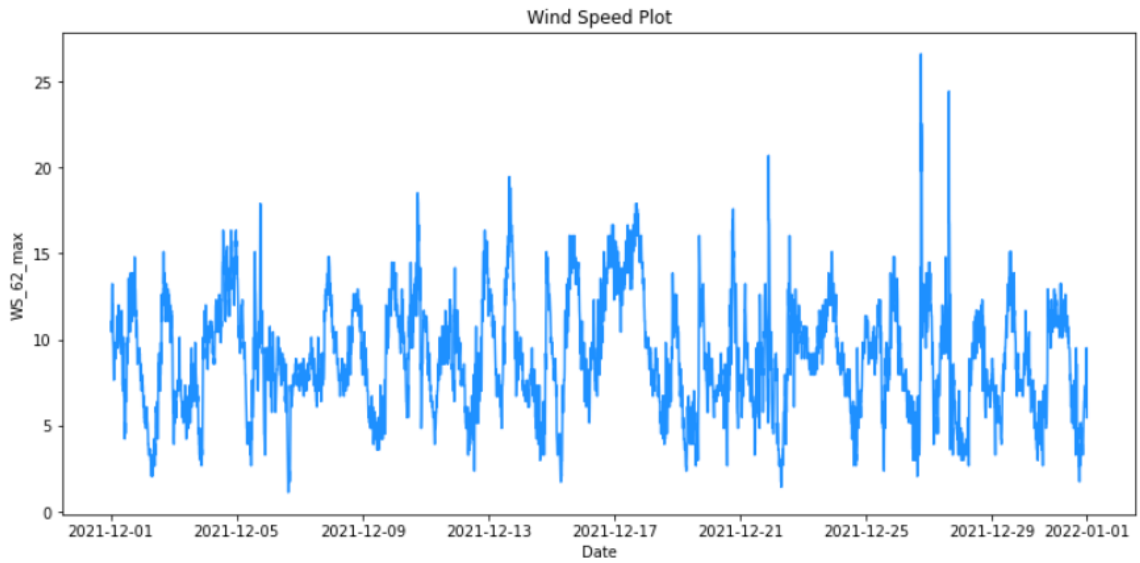

Fig. 3: Time series plot for the response (WS_62_max) feature.

Diagnostic plots for wind speed

The figures in Fig. 4 display diagnostic plots for the response variable. In the top left panel, a time series plot is depicted. The top right panel showcases a density plot indicating a right-skewed distribution with a longer tail on the right-hand side (RHS) compared to the left-hand side (LHS). Moving to the bottom left panel, a normal Q-Q plot is presented. Notably, the Q-Q plot points illustrate deviation from a straight line for higher endpoints, while the lower end adheres closely to a straight line. Additionally, the curve exhibits a significant tail to the right. Finally, the bottom right panel features a box plot where most points lie within the Interquartile Range (IQR) box, with only a few outliers falling outside the IQR.

3.3 Features selection results

Table 2 presents the feature importance analysis conducted using Lasso, which regulates model features by reducing certain regression coefficients to zero. The findings indicate that six of the 16 input features significantly forecast wind speed. These significant features include Pbaro_min, Pbaro_mean, Tgrad_min, Tgrad_max, Tgrad_stdv, and Pbaro_stv, as they possess notable regression coefficients. This alignment with scientific understanding is logical, as higher atmospheric pressure typically corresponds to increased wind speeds. Additionally, greater disparities in air temperature contribute to variations in atmospheric pressure, consequently influencing wind speed. These significant features will be utilised throughout the subsequent analysis to ensure the generation of accurate predictions.

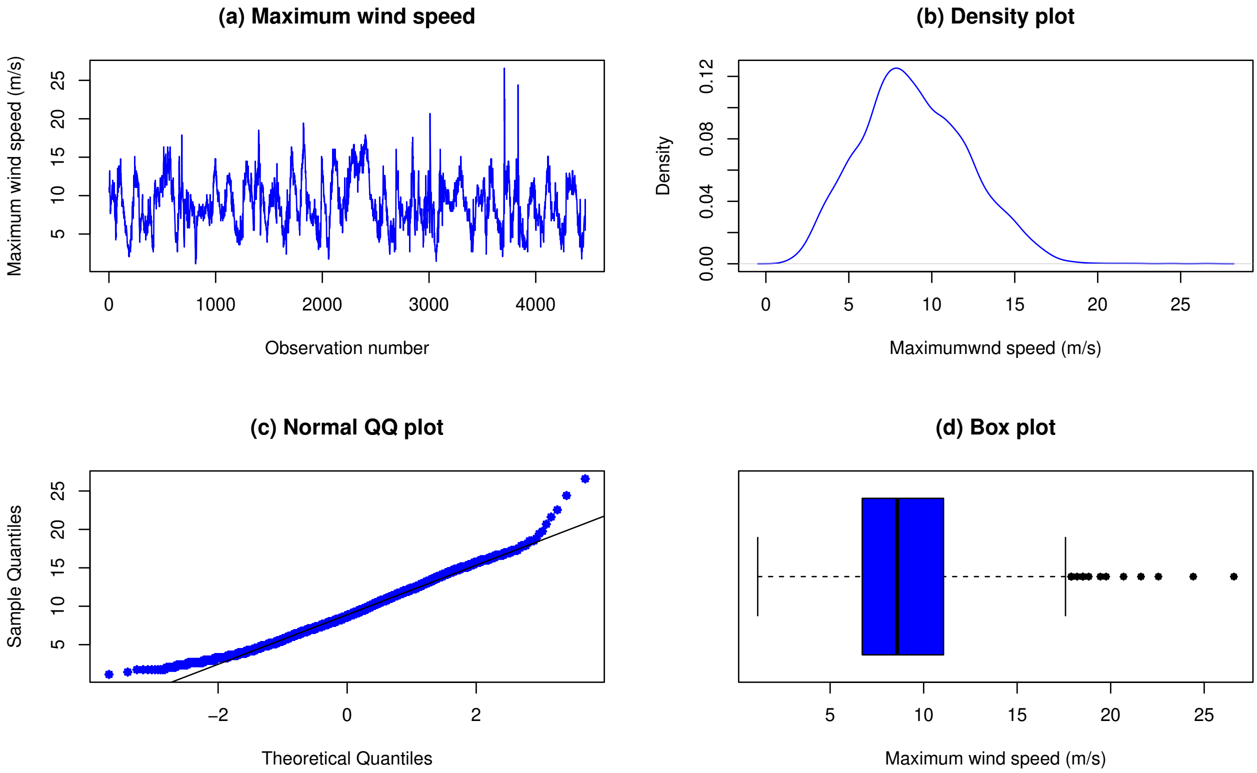

Fig. 4: Diagnostic plots for maximum wind speed at 62m (WS_62_max).

Table 2: Feature Importance.

Features	**Coefficient**
RH_min	0.003
RH_mean	0.042
RH_max	0.046
Tair_mean	0.070
RH_stdv	0.332
Tair_min	0.349
Tair_max	0.395
Pbaro_max	0.547
Tair_stdv	0.986
Tgrad_mean	1.052
Pbaro_min	3.960
Pbaro_mean	4.610
Tgrad_min	4.879
Tgrad_max	5.477
Tgrad_stdv	21.337
Pbaro_stdv	33.763

3.4 LSTM network results

Feature scaling ensures uniform scales across all variables before fitting the data to the LSTM network, enhancing the learning algorithm and expediting result generation. The dataset is partitioned into training and testing subsets in a 70% to 30% ratio. The LSTM network is configured with one feature, a hyperbolic tangent activation function, a batch size of one, 100 epochs, and a single output layer.

The predictions from both the training and testing sets are depicted in Figure 5. Predictions for the training set are highlighted in light grey, while those for the test set are in dark orange. Figure 5 illustrates the LSTM network's ability to generate accurate predictions, attributed to its capacity to capture longer-term trends.

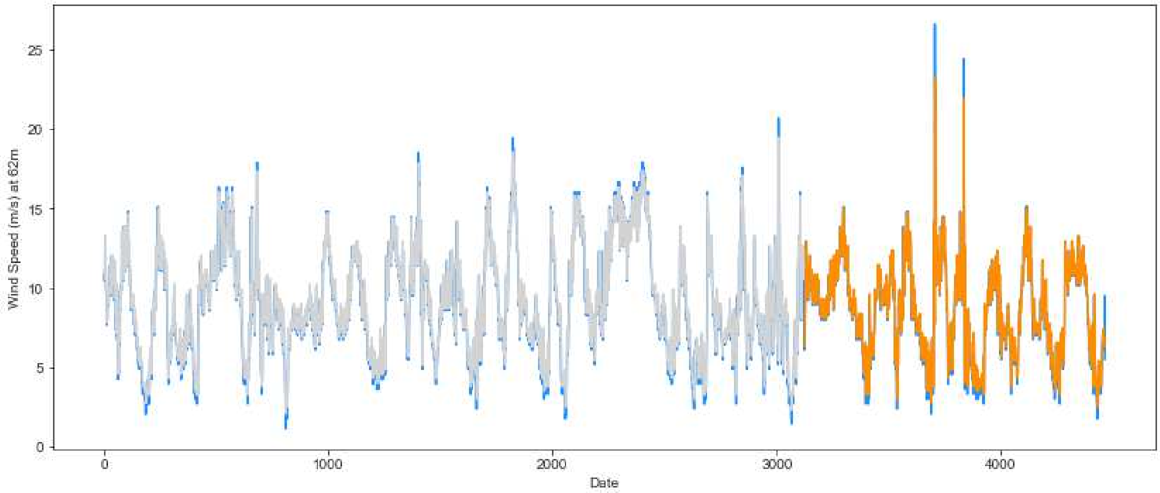

Fig. 5: Prediction for the full dataset.

The parameter setting for the LSTM network is given in Table 3.

Table 3: Parameter setting for LSTM network.

Parameters	Values
Number_features	1
Activation function	*tanh*
Batch_size	1
Epochs	100

3.5 SVR results

Various kernels are accessible, and it is crucial to identify the most suitable one for your dataset. In this study, the anticipation of wind speed at 62m above sea level in the western Cape Province is performed using three kernels: radial basis function, linear, and polynomial.

Figure 6 presents the results obtained using the radial basis function kernel, which was determined to be the most effective among the three kernels selected for the analysis. The results indicate that the fitted line fails to encompass the maximum points along the hyperplane, as indicated by the error margin (ε).

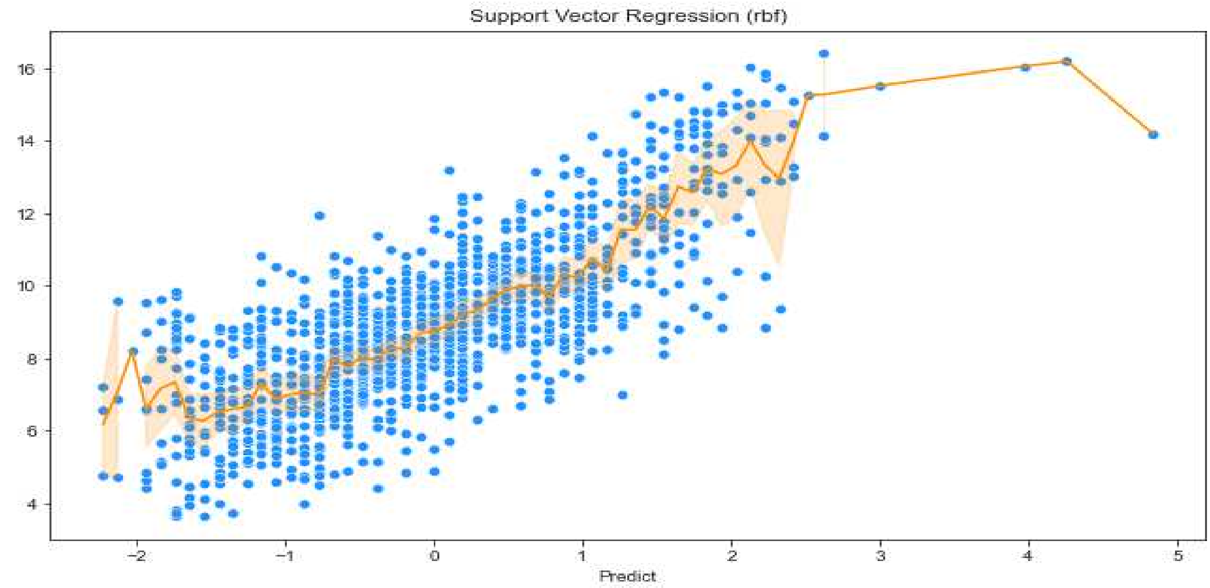

Fig. 6: Radial basis function plot for SVR.

3.6 Predictive skills and capabilities of the models

The predictive capabilities of the models were assessed using the Giacomini-White test, and the results are presented in Table 4.

We will use the following notation to indicate dominance, $M_j > M_i, \forall i \neq j$ to mean that M_j dominates M_i. From Table 4, for the prediction horizon $h = 224$, GPR $>$ SVR $>$ SGBM $>$ LSTM meaning that the GPR model has the highest predictive power since it dominates the other three models. Similarly, for prediction horizon $h = 670$, SVR $>$ SGBM $>$ LSTM $>$ GPR meaning the SVR model dominates the other three models.Likewise for $h = 1339$, SVR $>$ SGBM $>$ GPR $>$ LSTM implying the SVR dominates the other models.

Table 4: Model comparisons using the Giacommini-White test for different forecasting horizons.

h = 224			
Null hypothesis	Test statistic	p-value	Result
LSTM = SVR	9.6065	0.0226	Sign of the mean of the loss is (+) – SVR dominates
LSTM = SGBM	7.57912	0.7174	Sign of the mean of the loss is (+) – SGBM dominates
LSTM = GPR	10.9873	0.0041	Sign of the mean of the loss is (+) – GPR dominates
SVR = SGBM	2.11035	0.3481	Sign of the mean of the loss is (-) – SVR dominates
SVR = GPR	0.335857	0.8454	Sign of the mean of the loss is (+) – GPR dominates
SGBM = GPR	0.5696	0.7521	Sign of the mean of the loss is (+) – GPR dominates
h = 670			
Null hypothesis	Test statistic	p-value	Result
LSTM = SVR	32.192	0.0000	Sign of the mean of the loss is (+) – SVR dominates
LSTM = SGBM	13.7963	0.0010	Sign of the mean of the loss is (+) – SGBM dominates
LSTM = GPR	3.0270	0.2201	Sign of the mean of the loss is (-) – LSTM dominates
SVR = SGBM	17.0841	0.0002	Sign of the mean of the loss is (-) – SVR dominates
SVR = GPR	8.84261	0.0120	Sign of the mean of the loss is (-) – SVR dominates
SGBM = GPR	4.1218	0.1273	Sign of the mean of the loss is (-) – SGBM dominates
h = 1339			
Null hypothesis	Test statistic	p-value	Result
LSTM = SVR	174.882	0.0000	Sign of the mean of the loss is (+) – SVR dominates
LSTM = SGBM	244.546	0.0000	Sign of the mean of the loss is (+) – SGBM dominates
LSTM = GPR	23.3371	0.0000	Sign of the mean of the loss is (+) – GPR dominates
SVR = SGBM	13.9561	0.0009	Sign of the mean of the loss is (-) – SVR dominates
SVR = GPR	15.8329	0.0004	Sign of the mean of the loss is (-) – SVR dominates
SGBM = GPR	13.9445	0.0009	Sign of the mean of the loss is (-) – SGBM dominates

The predictive evaluation metrics for three forecasting horizons, $h = 224, 670$ and 1339, are given in Table 5. The results suggest that the GPR is the best-performing model for prediction horizon $h = 224$, while for horizons $h = 670$ and 1339, the SVR outperforms the other three models. These results are consistent with those from Table 4 on the comparative analysis of the predictive capabilities of the models using the Giacomini-White test.

Table 5: Predictive evaluation metrics for different forecasting horizons.

h = 224							
Model	RMSE	rRMSE	MAE	rMAE	MASE	MBE	PS(%)
LSTM	0.2137	0.0265	0.1669	0.0207	0.2274	-0.0114	-3.4
SVR	0.2221	0.0275	0.1644	0.0204	0.2241	0.0183	**-1.9**
SGBM	0.2210	0.0274	0.1650	0.0204	0.2249	0.0101	-2.3
GPR	**0.2071**	**0.0257**	**0.1613**	**0.0200**	**0.2199**	**0.0092**	
h = 670							
Model	RMSE	rRMSE	MAE	rMAE	MASE	MBE	PS(%)
LSTM	0.3161	0.0395	0.1863	0.0233	0.2515	-0.0260	8.5
SVR	**0.2519**	**0.0314**	**0.1453**	**0.0181**	**0.1962**	**0.0437**	**28.7**
SGBM	0.2531	0.0316	0.1643	0.0205	0.2218	-0.0053	19.3
GPR	0.6571	0.0820	0.2037	0.0254	0.2750	0.0200	
h = 1339							
Model	RMSE	rRMSE	MAE	rMAE	MASE	MBE	PS (%)
LSTM	0.3793	0.0443	0.2632	0.0307	0.3567	-0.1266	-11.2
SVR	**0.2239**	**0.0261**	**0.1478**	**0.0173**	**0.2003**	**0.0476**	**37.6**
SGBM	0.3638	0.0425	0.1768	0.0206	0.2396	0.0245	25.3
GPR	0.9015	0.1053	0.2368	0.0277	0.3209	0.0768	

Time series plots of maximum wind speed superimposed with predictions for the best-performing models for the forecasting horizons $h = 224, 670$ and 1339 are given in Figure 7.

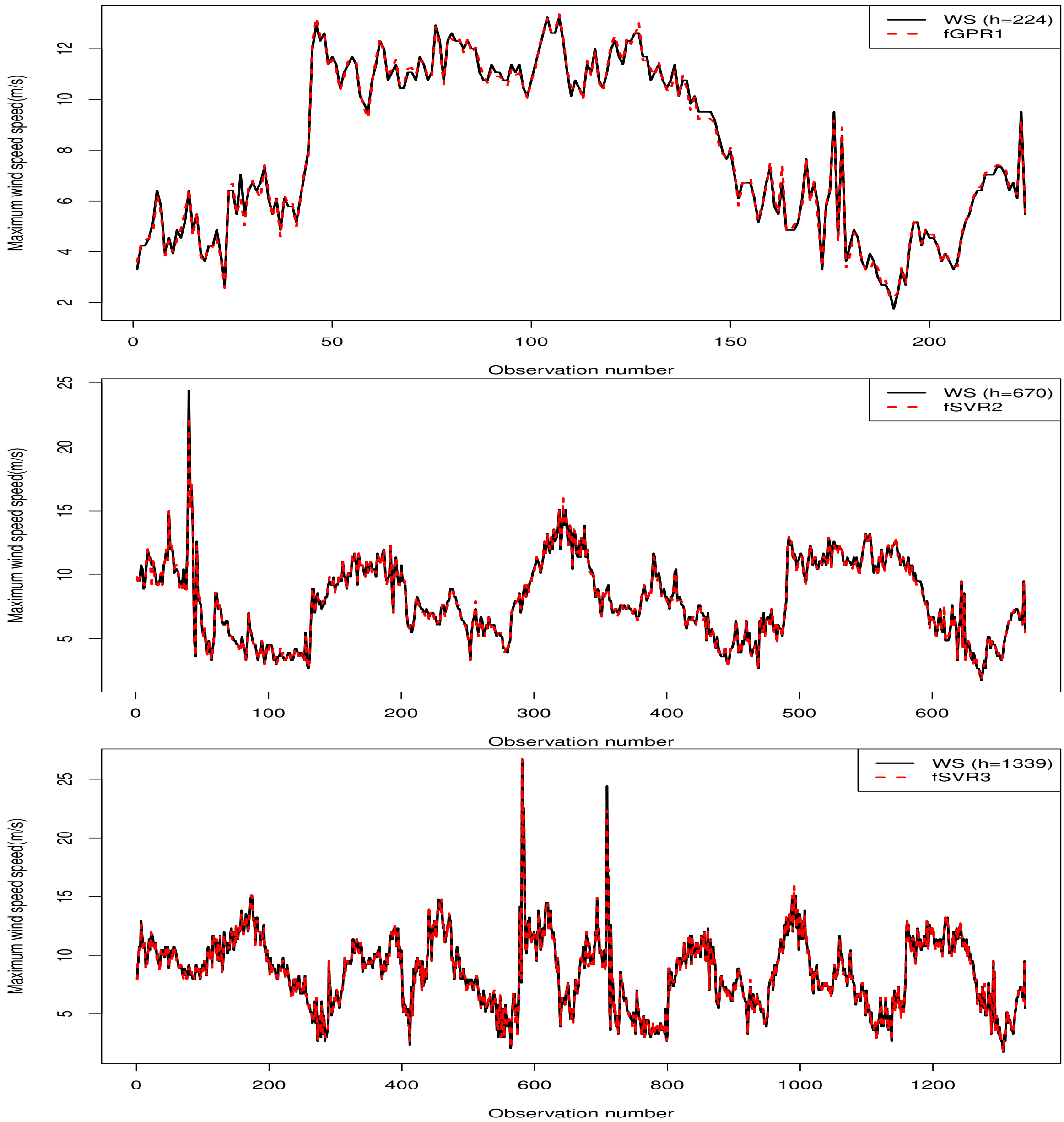

Fig. 7: Time series plots of maximum wind speed superimposed with predictions for the forecasting horizons $h = 224, 670$ and 1339.

4 Conclusion

In this study, we conducted a comparative analysis of several machine learning models. These models include long- and short-term memory, support vector regression, stochastic gradient boosting method, and Gaussian process regression. We used a dataset of wind speed measurements from Beaufort West meteorological station located in the Western province of South Africa to predict wind speed. Results showed that the support vector regression consistently outperformed the other models on different forecasting horizons, with the lowest mean absolute scaled error and the other evaluation metrics used. These findings are consistent with previous studies that have also shown the effectiveness of these models for predicting wind speed.

The results of this study have important implications for various applications that require accurate wind speed predictions, such as wind energy production, weather forecasting, and air quality monitoring. These applications can improve efficiency and effectiveness using machine learning models with high predictive accuracy, leading to significant economic and environmental benefits. However, it is important to note that the performance of these machine learning models is highly dependent on the specific dataset and the performance metrics used. Therefore, carefully selecting the appropriate model for a particular application and evaluating its performance using appropriate metrics is crucial.

In conclusion, the results of this study demonstrate the potential of machine learning models for predicting wind speed and provide useful insights for future research in this area. Further research is needed to evaluate the performance of these models under different weather conditions and to develop more accurate and reliable models for predicting wind speed.

Author Contributions:

Conceptualisation, KM and CS; methodology, KM and CS; software, KM and CS; validation, KM and CS; formal analysis, KM and CS; investigation, KM and CS; data curation, KM and CS; writing?review and editing, KM and CS; visualisation, KM; supervision, CS; project administration, CS. The contributions of these authors are equal in this work. All authors have reviewed and approved the final version of the manuscript for publication.

Funding:

The funding for this study was provided by the National e-Science Postgraduate Teaching and Training Platform (NEPTTP), administered by the Department of Science and Technology-Council for Scientific and Industrial Research (DST-CSIR)

Data Availability Statement:

The data was gathered from the USAid Venda station in Limpopo, South Africa. Access to minute-averaged intervals of the dataset was obtained on 13 August 2021 through the following link: http://wasadata.csir.co.za/wasa1/WASAData.

Conflicts of Interest:

The corresponding author declares that none of the authors have conflicts of interest.

References

[1] S. Sun, S. Wang, G. Zhang, J. Zheng. A decomposition-clustering-ensemble learning approach for solar radiation forecasting. *Solar Energy*, **163**:189–199, 2018. https://doi.org/10.1016/j.solener.2018.02.006

[2] N. Patel. Environmental and economic effects of fossil fuels. *J Recent Res Eng Tech*, **1**:1–12, 2014.

[3] S. Mishra, C. Bordin, K. Taharaguchi, I. Palu. Comparison of deep learning models for multivariate time series wind power generation and temperature prediction. *Energy Reports*, **6(3)**:273–286, 2020. https://doi.org/10.1016/j.egyr.2019.11.009

[4] M. Elsaraiti, A. Merabet. A Comparative Analysis of the ARIMA and LSTM Predictive Models and Their Effectiveness for Predicting Wind Speed. *Energies*, **14**, 6782, 1–16, 2021. https://doi.org/10.3390/en14206782

[5] R. Dhakal, A. Sedai, S. Pol, S. Parameswaran, A. Nejat, H. Moussa. A Novel Hybrid Method for Short?Term Wind Speed Prediction Based on Wind Probability Distribution Function and Machine Learning Models. *Applied Sciences*, **12**, 9038:1–19, 2022. https://doi.org/10.3390/app12189038

[6] T. Mutavhatsindi, C. Sigauke, R. Mbuvha. Forecasting hourly global horizontal solar irradiance in South Africa using machine learning models. *IEEE Access*, **8**:198872-198885, 2020. https://doi.org/10.1109/ACCESS.2020.3034690

[7] N.M. Odhiambo. Electricity consumption and economic growth in South Africa: A trivariate causality test. *Energy Economics*, **31(5)**:635–640, 2009. https://doi.org/10.1016/j.eneco.2009.01.005

[8] A. Alkesaiberi, F. Harrou, Y. Sun. Efficient Wind Power Prediction Using Machine Learning Methods: A Comparative Study. *Energies*, **15**, 2327:1–24, 2022. https://doi.org/10.3390/en15072327

[9] A.-N. Buturache, S. Stancu. Wind Energy Prediction Using Machine Learning. *Low Carbon Economy*, **12**:1–21, 2021. https://doi.org/10.4236/lce.2021.121001

[10] U. Singh, M. Rizwan. Analysis of wind turbine dataset and machine learning based forecasting in SCADA?system. *Journal of Ambient Intelligence and Humanized Computing*, **14**:8035–8044, 2023. https://doi.org/10.1007/s12652-022-03878-x

[11] Z. Tarek, M.Y. Shams, A.M. Elshewey, E.M. El-kenawy, A. Ibrahim, A.A. Abdelhamid, M.A. El-dosuky. Wind Power Prediction Based on Machine Learning and Deep Learning Models. *Computers, Materials & Continua*, **74(1)**:1–18, 2023. https://doi.org/10.32604/cmc.2023.032533

[12] L.O. Daniel, C. Sigauke, C. Chibaya, R. Mbuvha. Short-term wind speed forecasting using statistical and machine learning methods. *Algorithms*, **13(6)**, 132:1–30, 2020. https://doi.org/10.3390/a13060132

[13] H. Drucker, C.J. Burges, L. Kaufman, A.J. Smola, V. Vapnik. Support vector regression machines. *Advances in Neural Information Processing Systems*, **9**:155–161, 1996. https://proceedings.neurips.cc/paper_files/paper/1996/file/d38901788c533e8286cb6400b40b386d-Pap

[14] A. Zendehboudi, Z. Huan, C.C. Enweremadu. Evaluation of global solar radiation using multiple weather parameters as predictors for South African provinces. *Journal Cleaner Prod.*, **199**:272–285, 2018

[15] S. Hochreiter, J. Schmidhuber. Long short-term memory. *Neural Computation*, **9(8)**:1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[16] Y. Yu, X. Si, C. Hu, J. Zhang. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, **31(7)**: 1235–1270, 2019. https://doi.org/10.1162/neco_a_01199

[17] J.H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, **29**:1189–1232, 2001. https://doi.org/10.1214/aos/1013203451

[18] T. Hastie, R. Tibshirani, J. Friedman, J. Franklin. The elements of statistical learning: data mining, inference and prediction. *Math. Intell.*, **27**, 83–85, 2005.

[19] J. Bien, J. Taylor, R. Tibshirani. A lasso for hierarchical interactions. *Annals of Statistics*, **41(3)**:1111–1141., 2013. https://doi.org/10.1214%2F13-AOS1096

[20] R. Tibshirani. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58(1)**:267?288, 1996. https://www.jstor.org/stable/2346178

[21] R. Giacomini, H. White. Tests of conditional predictive ability. *Econometrica*, **74(6)**, 1545-1578, 2006. https://doi.org/10.1111/j.1468-0262.2006.00718.x

[22] J. Lago, G. Marcjaszd, B. De Schuttera, R. Weron. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, **293**, 116983:1–21. 2021. https://doi.org/10.1016/j.apenergy.2021.116983

[23] R. Ihaka, R. Gentleman. R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, **5(3)**, 299-314, 1996. https://doi.org/10.2307/1390807

[24] Team, RC, 1999. Writing R extensions. Version 4.3.3. R Foundation for Statistical Computing. Available online (Accessed on 27 September 2023) https://cran.r-project.org/doc/manuals/R-exts.html

[25] G. Van Rossum. Python Programming Language. *In USENIX annual technical conference*, **41(1)**:1–36), 1989.