Original Research Article

Analysis and Forecasting of Solar Power Generation using Machine Learning

Todizara Andrianajaiana[1], Tsivalalaina David Razafimahefa[2], Haba Cristian-Gyozo[3], Dorin Dumitru Lucache[4]*

1 Doctoral school EDT-ENRE, University of Antsiranana, Antsiranana, Madagascar, david.razafimahefa@gmail.com

2 Doctoral school EDT-ENRE, University of Antsiranana, Antsiranana, Madagascar, todizara.andrianajaina@gmail.com

3 Faculty of Electrical Engineering, Technical University of Iasi, Iasi, Romania, cghaba@tuiasi.ro

4 Faculty of Electrical Engineering, Technical University of Iasi, Iasi, Romania, dorin.lucache@tuiasi.ro

***Corresponding author**： *todizara.andrianajaina@gmail.com*

Abstract: Current photovoltaic systems are equipped with a monitoring system. The data is recorded with a predefined time base. The operator has a large amount of data. This article is a contribution to the analysis and understanding of this data. It proposes to estimate Generation from this data using machine learning. The study was conducted at the 1Mw generation site in Miroslava, Iasi, Romania. The visualization and interpretation of the generation data is presented in this article. Then, prediction techniques are presented to obtain an estimate of the plant's generation. These techniques are: Simple Exponential Smoothing (SES), Autoregression (AR), Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors (SARIMA), Holt Winter's Exponential Smoothing (HWES), Long Short Term Memory neural (LSTM) and Convolutional Neural Network (CNN).

Keywords: Power generation, Artificial inteligence, Fore-casting, Solar power generation, Data processing

1. Introduction

The growing demand for accurate energy generation fore-casts within the renewable energy sector, as highlighted by Khatibi et al. (2022) in their recent research[1], can be attributed to several driving factors. These include the diminishing support systems for traditional energy sources and the evolving dynamics of the energy market, which present challenges and opportunities for energy operators and stakeholders. Furthermore, the ongoing integration of renewable energy sources into the broader energy landscape requires seamless coordination with grid operations and the efficient balance of supply and demand. As a result, generation forecasting has emerged as a crucial bridge spanning diverse temporal horizons, catering to the multifaceted needs of various stake-holders.

It is essential to acknowledge that renewable energy sources, such as solar power, are inherently susceptible to the whims of weather conditions, making them inherently challenging to predict and schedule. The inherent stochastic nature of energy generation from these sources compounds the intricacies of forecasting. In response to this challenge, advanced technologies for renewable energy forecasting and scheduling have been developed and implemented. Predicting energy generation goes beyond mere anticipation; it serves as a strategic tool for utilities to optimize resource allocation and generation processes.

In the contemporary era, significant advancements in artificial intelligence have revitalized energy forecasting, elevating it to the status of a predictive science. Through the utilization of estimation algorithms and the power of machine learning, software systems have gained the capability to discern patterns in current data

Received: 9 June 2023 Accepted: 25 August 2023 Available online: 15 Novermber 2023

and conduct sophisticated analyses to predict the future with a high degree of accuracy. The foundation of this endeavor lies in the creation of tailored forecasting models, finely attuned to specific contexts over varying timeframes, enabling precise predictions of energy generation. However, the effectiveness of these models is inherently tied to the quality of the underlying data, which must be valid, accurate, reliable, consistent, and complete[2,3].

This paper places a strong emphasis on the pivotal role of artificial intelligence and machine learning in addressing the complexities of energy forecasting. It is dedicated to providing acomprehensive exploration of generation data, using the Miroslava site in Iasi, Romania, as an illustrative case. In this study, meticulous analysis was performed on data collected at hourly intervals over a two-year span. Subsequently, the paper delves into the heart of the matter: conducting comparative analyses of various forecasting methodologies used to predict energy generation at the Miroslava site. As such, the primary contribution of this paper lies in its comparative analysis of diverse forecasting methods, encompassing SES, AR, SARIMA, HWES, LSTM, and CNN, all aimed at predicting energy generation at the Miroslava site.

2. Materials and Methods

2.1 The study site

2.1.1 Romanian Energy Situation

Romania might be entirely self-sufficient in terms of energy. The overall output of all power-producing facilities is 62 billion kWh, or 124% of its own requirements. Despite this, Romania trades electricity with other countries. The entire electrical energy usage in the country is 49.64 billion kWh per year. This amounts to an average use of around 2,597 kWh per resident per year[4–6].

Romania uses a variety of energy sources. These sources contribute to the total output. The country's generation statistics are provided in real-time on the website of Transelectrica, the national energy transport corporation.

The Figure 1 below displays some measurement data for October 30, 2022[7].

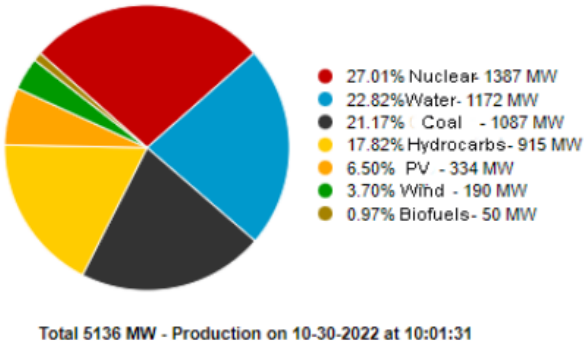

Figure 1 Romanian energy generation on October 30,2022 at 10h01 PM.

The Table I shows Romania's total output for the month of October 28, 2022[7].

Table 1 Romanian Energy Generation on October 28,2022.

Energy sources	Generation [Mw]
Coal	1481
Hydrocarbs	1499
Nuclear	1388
Wind	127
Water	1379
Biofuels	71
Photovoltaics	178

Solar photovoltaics play an important role in the country's energy mix. The Miroslava site in the Modova area is a participant in the energy generation industry. Moldova is one of Romania's coldest areas. July has the greatest daily temperatures in Moldova, with an average of 29.4 °C. The coldest month, on the other side, is January, with an average temperature of only 1.8 degrees Celsius.

2.1.2 Miroslava Power Plant

The study site and photovoltaic park in Ciurbes¸ti village, Miroslava. It is the largest and modern photovoltaic power plant in Iasi county. The park is located on the outskirts of the city of Iasi on an area of 5 hectares and is operated with tracking systems. The tracking rotate according to the sun so that they can capture solar energy as efficiently as possible. With a capacity of about 1 megawatt, the plant provides the community with energy.

In addition, the administration of the City Hall of Miroslava expects additional revenue in the budget, which will allow to finance other projects aimed at developing the district by becoming a producer of electricity. Miroslava has become an electricity producer. The park provides the energy needed for the public facilities of the municipality and for public lighting. The 4 200 photovoltaic panels are mounted on 97 trackers that follow the movement of the sun. Figure 2 illustrates the site's appearance.

Figure 2 Miroslava PV power plant.

The Ciurbes¸ti photovoltaic park has been producing electricity since the end of July 2015. With the commissioning of the photovoltaic plant, Miroslava City Hall has practically become an electricity producer, with part of the generation being fed into the national energy system. Together with the ecostamp levies, this brings new financial resources to the municipal budget.

2.1.3 Data Used

In hourly intervals, the site logs all generation data and meteorological factors (irradiance, temperature, wind speed). We used data from the previous two years for the study (2021-2022). As a result, these data comprise 19728 records. The Figure 3 summarizes these findings.

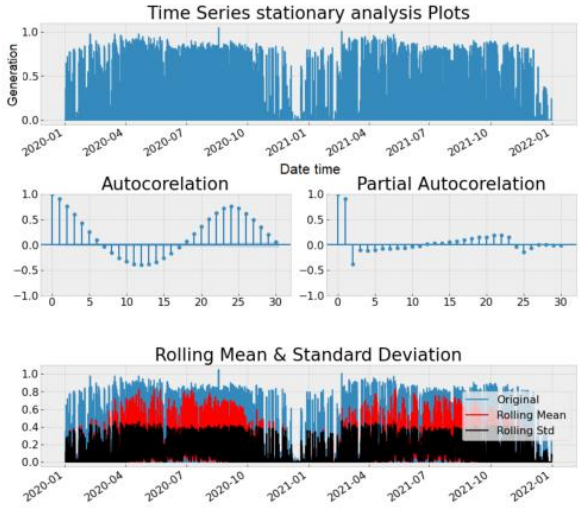

Figure 3 Solar irradiance data summarize.

It should be emphasized that the Rolling mean is a statistical average that is used to examine ordered series of data, such as the time series presented here, by reducing transitory oscillations to highlight longer-term patterns. This average is considered to be moving because it is recalculated continually, utilizing a subset of items at each computation in which a new element replaces or is added to the subset. A rolling mean assists us in detecting trends that might otherwise be difficult to discern[8,9].The mathematical formula for calculating it is as follows:

$$\overline{x_n} = \frac{1}{N}\sum_{k=0}^{N-1} x_n - k \quad \text{(Eq.1)}$$

Astandard deviation measures how far the data deviate significantly from the mean. A low standard deviation indicates that data is clustered around the mean, whereas a high standard deviation indicates that data is more dispersed. A standard deviation close to zero implies that the data sets are close to the mean, while a low or high standard deviation indicates that the sets of values are above or below the mean respectively[10]. The mathematical formula for calculating it is as follows:

$$S_n = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x_n}\right)} \quad \text{(Eq.2)}$$

2.2 Machine Learning and Forecasting Method

Artificial intelligence is defined in a variety of ways. Artificial intelligence, in its most basic form, referring to the reproduction of human intelligence in computers that are designed to think like humans and copy their activity patterns. Artificial intelligence, in other terms, is an area of computer science that allows a computer system to simulate human intellect. We anticipate that machine learning (ML) and deep learning (DL) will be explored as part of Artificial Intelligence.

This article delves further into machine learning[11]. The learning methods employed will be discussed more below.

2.2.1 Simple Exponential Smoothing (SES)

The SES technique models the future time step as an exponentially weighted linear function of data from previous time steps. In order to work well, this approach requires our time series to be nonstationary (no trend or seasonality)[12,13]. It can be written as:

$$y_{t+1} = \alpha y_t + (1-\alpha) yt \quad \text{(Eq.3)}$$

where $0 \le \alpha \le 1$ denotes the smoothing parameter. The forecast for time t+1 is a weighted average of all observations in the series y1,...,yt. The parameter α controls the rate at which the weights decline. If the smoothing parameter is close to zero, the prior fitted value *hatyt* is given more weight and the new information is ignored. The forecaster can determine how to estimate the data and filter out noise by modifying the smoothing parameter value.

2.2.2 Holt Winter's Exponential Smoothing (HWES)

Triple Exponential Smoothing, sometimes referred to as HWES. The Holt-Winters approach encodes multiple historical values with exponential smoothing and utilizes them to forecast "typical" values for the present and future. It is a method for smoothing the results of a Univariate Time Series Analysis so that they may be used to anticipate future values. The concept is to assign exponentially diminishing weights to more recent instances. As we proceed back in time, we will witness decreasing weights[14,15].

The seasonality component, st,for m seasons per period, is an extension of Holt's technique in the Holt-Winters method. The multiplicative and additive variants of this model are also available. According to the additive technique, the seasonal component accumulates to about zero throughout the year and the error variance is assumed to be constant. The multiplicative version makes the assumption that the error variance scales with the level and that the annual total of the seasonal component is about m .

The multiplicative Holt-Winters method is given as:

$$y_{t+h} = (l_t + hb_t)s_{t+h-m(k+1)} \quad \text{(Eq.4)}$$

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \quad \text{(Eq.5)}$$

$$b_t = \beta^* \left(l_t - l_{t-1} + (1-\beta^*)b_{t-1}\right) \quad \text{(Eq.6)}$$

$$s_t = \gamma \frac{y_t}{l_{t-1} - b_{t-1}} + (1-\gamma)s_{t-m} \quad \text{(Eq.7)}$$

$s_{t+h-m(k+1)}$ is always predicated on the previous seasonal period since k is the modulus of $(h-1)/m$. It is a weighted average (α weighting) between the non-seasonal projection and the seasonally adjusted observation. The trend component remains the same. A weighted average (γ weighting) between the present seasonal index and the same season of the previous season period makes up the seasonal component.

2.2.3 Autoregression (AR)

The autoregression approach (AR) predicts the next step in the sequence as a linear function of previous time steps' data. Model parameters include: The number of AR (autoregressive) terms (p): p is the parameter connected with the model's autoregressive feature, which includes previous values, i.e. lags of the dependent variable[16].

When a time series value is regressed on earlier values from the same time series, this is known as an autoregressive model. An illustration is y_t on y_{t-1} given as:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \quad \text{(Eq.8)}$$

The previous time period's response variable has been transformed into the predictor in this regression model, and the errors follow the standard assumptions for errors in a basic linear regression model. The number of values in the series that came before the current value were utilized to forecast it, which is known as the order of an autoregression. Thus, the model that came before is a first-order autoregression, abbreviated as AR(1).

The autoregressive model for predicting y this year (y_t) using measurements in the two years prior (y_{t-1}, y_{t-2}) is as follows:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t \quad \text{(Eq.9)}$$

Due to the fact that the value at time t may be predicted given the values at periods $t-1$ and $t-2$, this model is a second-order autoregression, abbreviated as AR(2). The value of the series at each time t is a function of the values at periods $t-1, t-2, \cdots t-k$ in a multiple linear regression, which is known as a k^{th} -order autoregression or AR(k).

2.2.4 Seasonal Auto-Regressive Integrated Moving Average with eXogenous Factors (SARIMAX)

An SARIMAX model is, in essence, a linear regression model that employs a Seasonal Auto-Regressive Integrated Moving Average type process. This model is beneficial when we think that the residuals have a seasonal trend or pattern[17].

$$\begin{aligned} w_t &= y_t - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \cdots - \beta_b x_{b,t} \left(1 - \sum_{i=1}^{p} \phi_i L^i\right)\left(1 - \sum_{j=1}^{p} \phi_i L^{j\times 8}\right)(1-L)^d \left(1-L^8\right)^D \\ w_t - \eta &= \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\left(1 + \sum_{j=1}^{Q} \theta_j L^{j\times 8}\right)\alpha_t, \alpha_t \sim i.i.d \sim \Phi\left(0, \sigma^2\right) \end{aligned} \quad \text{(Eq.10)}$$

Where:

- L : lag operator.
- y_t :observed output at t.
- $x_{k,t}$: $k-th$ exogenous input variable at t.
- β_k :coefficient for the $k-th$ explanatory input.

- b :number of exogenous input.
- w_t :auto-correlated regression residuals.
- p :non-seasonal AR component's order.
- P :seasonal AR component's order.
- q :non-seasonal MA component's order.
- Q :seasonal MA component's order.
- s :seasonal length.
- D :time series's seasonal integration order.
- η :constant in the SARIMA model
- α_t :innovation, shock or error term at time t.
- α_t :time series observations are independent and identically distributed (i.e. i.i.d) and follow a Gaussian distribution (i.e. $\Phi\left(0,\sigma^2\right)$)

Assuming that the differences (both seasonal and non-seasonal) result in a stationary time series after reordering the parameters in the formula 10 above (z_t)yields:

$$\begin{aligned} z_t &= (1-L)^d \left(1-L^s\right)^D w_t \\ \mu = E[z_t] &= \frac{\eta}{\left(1-\phi_1-\phi_2-\cdots-\phi_p\right)\left(1-\Phi_1-\Phi_2-\cdots-\Phi_p\right)}\left(1-\sum_{i=1}^{p}\phi_i L^i\right)\left(1-\sum_{j=1}^{p}\Phi_j L^{j\times 8}\right)(1-L)^d\left(1-L^s\right)^D \\ \left(w_t-\mu\right) &= \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\left(1+\sum_{j=1}^{Q}\Theta_j L^{j\times 8}\right)\alpha_t \end{aligned} \qquad \text{(Eq.11)}$$

2.2.5 Long Short-Term Memory (LSTM)

Artificial neural networks with Long Short-Term Memory (LSTM) are employed in deep learning and artificial intelligence. LSTMs, unlike traditional feedforward neural networks, feature feedback connections. This form of recurrent neural network (RNN) can analyze complete data sequences as well as individual data points. A conventional RNN is described as having both "long-term memory" and "short-term memory" in the moniker of LSTM. The network's activation patterns change once each time step, comparable to how physiological changes in synaptic strength store short-term memory. The network's linking weights and biases change once every training session, analogous to how physiological fluctuations in synapse strength store long-term memory. The term "long short-term memory" refers to the short-term memory that the LSTM architecture aims to provide RNN with, which can withstand thousands of timesteps.[18,19]

The lowercase variables in the following equations denote vectors. The matrices W_q and U_q include, respectively, the weights for the input and recurrent connections, where the subscript q may refer to the input gate i, the output gate o, the forgetting gate f, or the memory cell c, depending on the activation to be computed. So, in this part we will use the "vector notation" So, for example, $c_t \in \mathbb{R}^h$ is not only a unit of one LSTM cell, but contains units of h LSTM cells.

$$\begin{aligned} f_t &= \sigma_g\left(W_f x_t + U_f h_{t-1} + b_f\right) \\ i_t &= \sigma_g\left(W_i x_t + U_i h_{t-1} + b_i\right) \\ o_t &= \sigma_g\left(W_o x_t + U_o h_{t-1} + b_o\right) \\ c_t &= \sigma_c\left(W_c x_t + U_c h_{t-1} + b_c\right) \\ c_t &= f_t \Theta c_{t-1} + i_t \Theta c_t \end{aligned} \qquad \text{(Eq.12)}$$

where the initial values are $c_o = 0$ and $h_o = 0$ and Θ Hadamard product is indicated by the operator. The time step is indicated by the subscript t.

Variables:

- $x_t \in \mathbb{R}^d$:input vector to the LSTM unit.
- $f_t \in (0,1)^h$:forget gate's activation vector.
- $i_t \in (0,1)^h$:input/update gate's activation vector.

- $o_t \in (0,1)^h$:output gate’ s activation vector.
- $h_t \in (-1,1)^h$:hidden state vector also known as output vector of the LSTM unit.
- $\tilde{c}_t \in (-1,1)^h$:cell input activation vector.
- $c_t \in \mathbb{R}^h$:cell state vector.
- $W \in \mathbb{R}^{h \times d}, U \in \mathbb{R}^{h \times h}$ and $b \in \mathbb{R}^h$:weight matrices and bias vector parameters which need to be learned during training.

where the superscripts d and h refer to the number of input features and number of hidden units, respectively.

Activation functions:

- σ_g :[[sigmoid function]].
- σ_c :[[hyperbolic tangent]] function.
- σ_h :hyperbolic tangent function or, as the peephole LSTM , $\sigma_h(x) = x$.

2.2.6 Convolutional Neural Network (CNN)

ACNN or convnet is a machine learning subset. It is one of several types of artificial neural networks utilized for diverse applications and data sources. An input layer, hidden layers, and an output layer make up a convolutional neural network. The middle layers in any feed-forward neural network are known as hidden layers because the final convolution and the activation function hide their inputs and outputs. The hidden layers of a convolutional neural network contain convolutional layers. This typically contains a layer that creates a dot product of the input matrix of the layer and the convolution kernel. ReLU often serves as the activation function for this product, which is typically the inner Frobenius product. The convolution procedure develops a feature map as the convolution kernel moves over the input matrix for the layer, adding to the input of the following layer. Following this are further layers like normalizing, pooling, and fully linked layers[20].

3. Results and Discussion

3.1. Generation Analysis

Consider the information in Section II-A3.To get a better view of our variable, or to be able to see the dispersion or centrality of the distribution of these values, we will present them in a boxplot.

It should be noted that this data indicates that the outliers are the generations exceeding 0.752MW. The maximal generation is 1.047MW. Only 3.42% of output hours have achieved a reasonably high level in the past two years. 50% of the output is between 0.084MW and 0.721MW. The average output is 0.405MW.

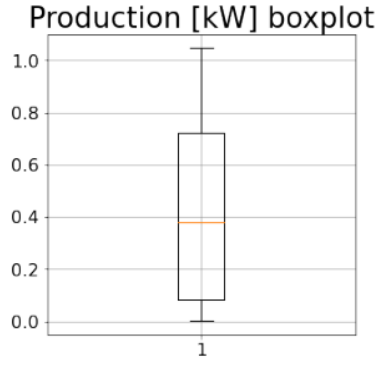

Figure 4 Generation boxplot.

Romania, has a continental climate with hot summers and harsh winters with heavy rain in the mountains. This global trend is, however, nuanced by relief and a slight Mediterranean influence to the southeast. In the summer, the average temperature ranges between 22 and 24 degrees Celsius. The maximum temperature can occasionally reach 38 ℃. The average temperature in winter is -3 ℃[21–23]. In the case of Iasi, the seasons are as follows:

- **Spring**: March 21 to June 20
- **Summer**: June 21 to September 20
- **Autumn**: September 21 to December 20
- **Winter**: December 21 to March 20

The Figure 5 represent the global temperature:

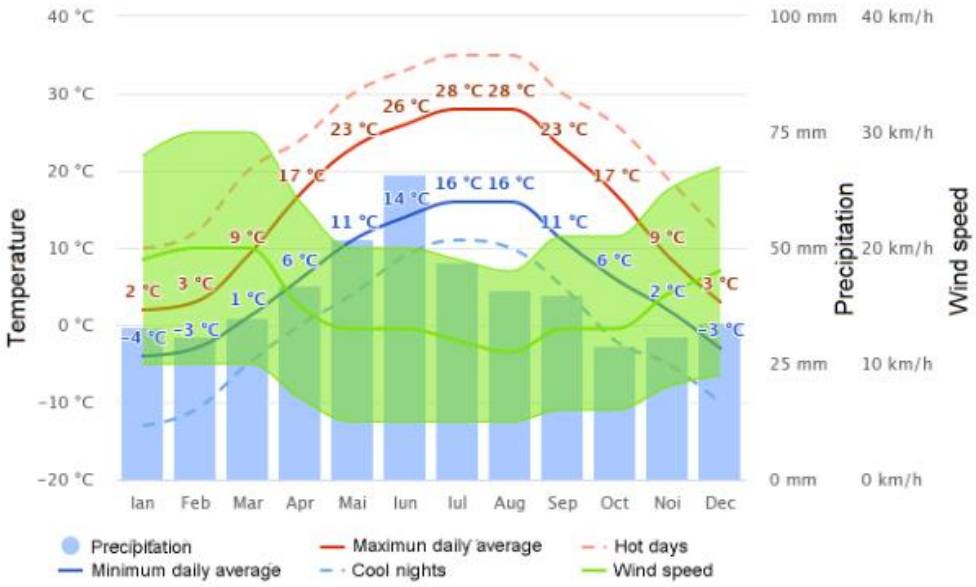

Figure 5 Average temperature, precipitation and wind speed in Iasi.

The mean daily maximum (solid red line) displays the typical high temperature for Iasi for one day in each month. The average minimum temperature is also shown by the “average daily minimum” (solid blue line). The average of the warmest day and coldest night of each month over the past 30 years is represented by the dotted blue and red lines. You can plan for average temperatures and be ready for days when it’s warmer or colder. The green area indicates the speed of the wind.

The Figure 6 below depicts Iasi global solar irradiance.

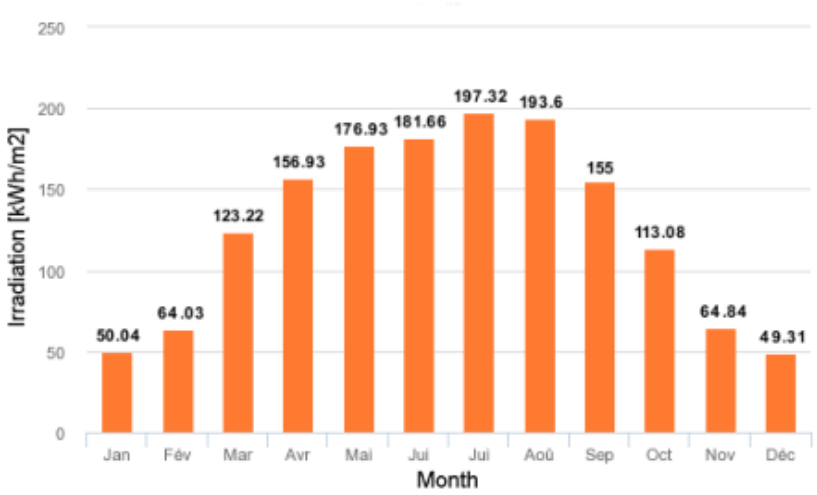

Figure 6 Monthly irradiation on a fixed plan.

These parameters influence the global generations. The following Figure 8 shows the profile of the typical day generation, witch is the hourly average for each day of the week.

The average generation for each month of the year is displayed in Figure 7 below.

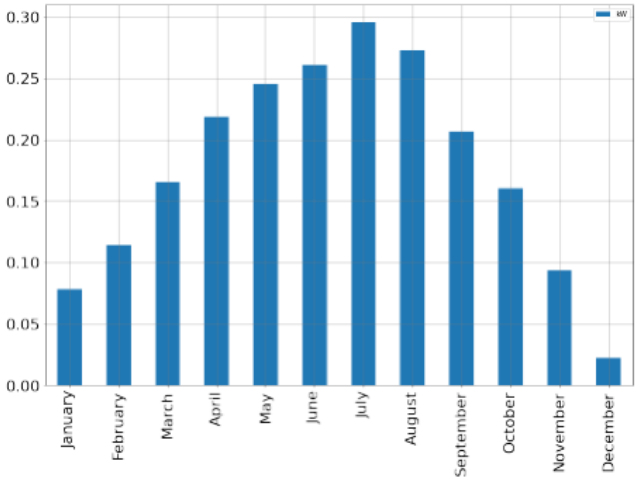

Figure 7 Monthly generation average profile.

We see that the highest output occurs in July and August. Less is in March and December.

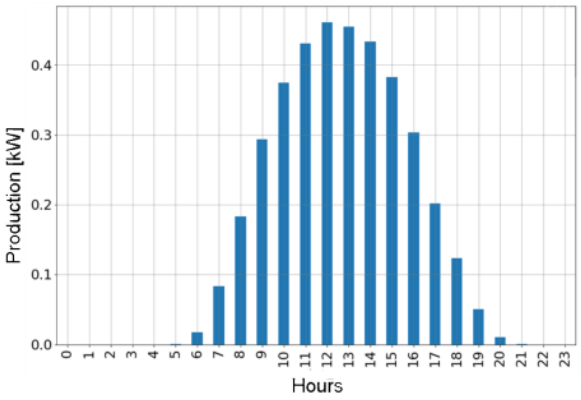

Figure 8 Daily generation profile.

If we take the daily average of total generation, we get the day week profile, show on Figure 9:

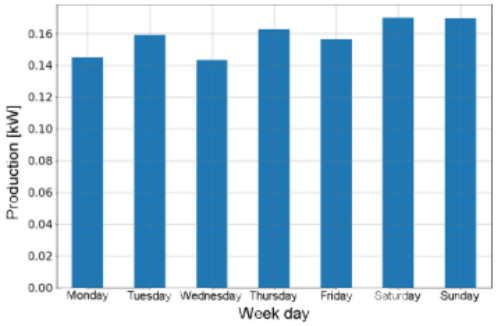

Figure 9 Weekly generation profile.

We can see that generation is a little bit lower at the start of the week than it is for the entire week. In contrast, the weekend has the highest of any day of the week. This is probably because regular cleaning and maintenance are done every weekend. The energy demand of the distribution, which is noticeably higher at the start of the week, is another factor.

3.2 Generation Forecasting

Six learning algorithms that we created and put into use can estimate the value of generation. SES, WHES, AR, SARIMAX, LSTM, and CNN are the algorithms in question. We divided the data into three categories for

each algorithm: learning, learning validation, and estimation tests. These categories are as follows:

- 19201 hours of training data
- 264 hours of validation data
- 263 hours of testing data

The corresponding outcome of these algorithms estimation is shown in the following Figure 10.

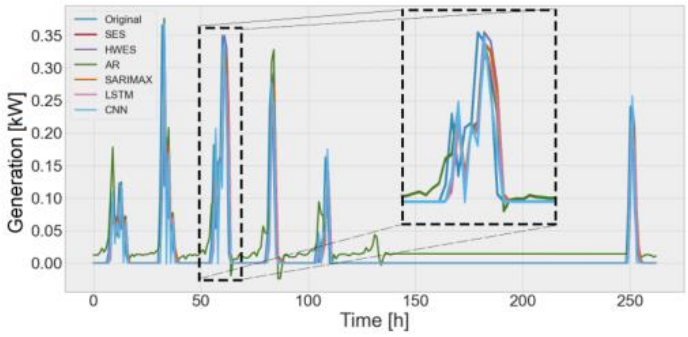

Figure 10 Generation Forecasted.

We used the following metrics to rate the accuracy of each algorithm's prediction:

• Mean Absolute Error (MAE): [24–26]Without taking into account their direction, MAE calculates the average size of mistakes in a set of forecasts. All individual differences are equally weighted in the test sample's average of the absolute disparities between prediction and observation.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i\right| \quad \text{(Eq.13)}$$

• Root mean squared error (RMSE): [24],[26],[27] It is a scoring formula for quadratic equations that also calculates the average error magnitude. It is the average of the squared discrepancies between predicted results and actual observations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i\right)^2} \quad \text{(Eq.14)}$$

• Mean absolute percentage error (MAPE): [25],[26],[28]it is often referred to as the mean absolute percentage deviation (MAPD), is a metric for forecasting technique accuracy. The accuracy is often expressed as a ratio determined by the following formula:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i}{y_i}\right| \quad \text{(Eq.15)}$$

• Coefficient of determination R^2: [26],[29] The percentage of variance in the dependent variable that can be predicted from the independent variables is known as the coefficient of determination, abbreviated R^2 and pronounced "R squared". It is a statistic applied to statistical models whose main objective is to either forecast future events or test hypotheses using data from other relevant sources. Based on the percentage of overall variance in outcomes that the model accounts for, it gives a gauge of how well-observed results are duplicated by the model.

A set of data has n labeled values. $y_1,\cdots,y_n$, each with a predetermined value $y_1,\cdots,y_n$.

Residuals are stated as $e_i = y_i - y_i$.

Considering that $\bar{y}$ is the observed data's mean:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{(Eq.16)}$$

then the variability of the data set can be measured with two Mean squared error formulas:

- The residual sum of squares (also known as the mean square error)

$$SS_{res} = \sum_i \left(y_i - y_i \right)^2 = \sum_i e_i^2 \quad \text{(Eq.17)}$$

- The total sum of squares:

$$SS_{tot} = \sum_i \left(y_i - \bar{y} \right)^2 \quad \text{(Eq.18)}$$

The most general definition of the coefficient of determination is

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad \text{(Eq.19)}$$

In the ideal scenario, the predicted values and observed values coincide precisely, which leads to $SS_{res} = 0$ and $R^2 = 1$. A baseline model, which always predicts $\bar{y}$, will have $R^2 = 0$. Models that have worse predictions than this baseline will have a negative R^2.

Figure 11 shows us the result for each metric, for each algorithm.

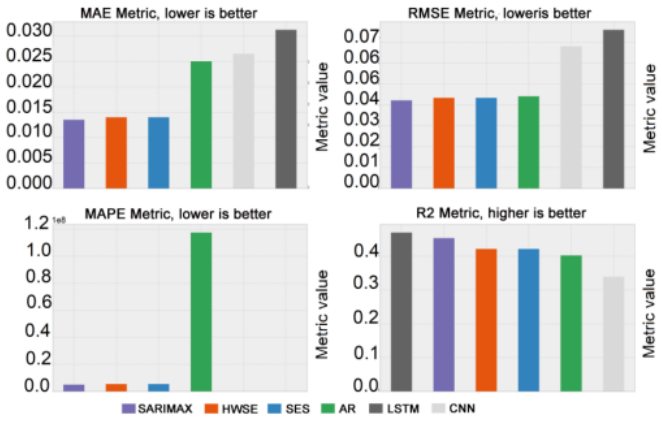

Figure 11 Metrics for each algorithm.

Figure 12 displays the boxplot of the residuals to help us better understand the accuracy of the estimate.

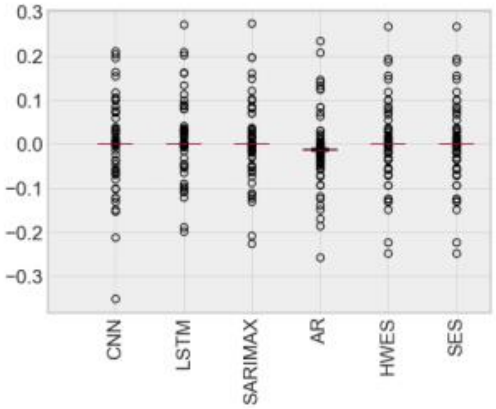

Figure 12 Metrics for each algorithm.

The residuals are all centered on zero, except for the case of the autoregression whose mean of the residuals is -0.01. These outcomes enable us to come to the conclusion that the site Generation forecasting techniques employed are reliable. But for generation forecasting, the LSTM neural network performs the best.

4. Conclusion

To summarise, the examination of the output of an energy-producing solar site is the main focus of this article. This is where Miroslav, Iasi, Romania, is located. The location has a 1Mw installed power. The site data from January 1, 2020, to December 31, 2021, were the data used in this investigation. We observed that the site operates at its peak efficiency on 3.42 of the data collected over the course of these two years, or 675 hours, which still equals 28 days cumulatively. The output is 0.405Mw on average.

The months of May, June, July, and August, or the summer, have greater output rates. It is unmistakably low for winter season, specifically December and January. In terms of the daily profile, output starts at 5 a.m., peaks between 12 and 1p.m., and starts to go down starting at 9 p.m. We've also noted that productivity is higher on weekends, which might be attributed to weekend maintenance and cleaning activities.

Before using learning techniques to make forecasts, we preprocessed the data. After that , the data is split into three sets: one for learning, one for validation, and one for testing the prediction's accuracy. Six techniques, including SES, WHES, AR, SARIMAX, LSTM, and CNN, were evaluated. To evaluate each method's quality in terms of forcating, we looked at its metrics. All of the approaches had positive results, meaning that the forecasted values were quite close to the initial value that was recorded. All of the residuals have a zero center. However, SARIMAX and neural networks, specifically LSTM and CNN, produced the greatest results of all the techniques examined in this study.

Conflict of Interest

The authors declare no conflict of interest.

References

1. A. Khatibi, M. H. Jahangir, F. Razi Astaraei, and F. Mohabbati, Predicting the renewable energy consumption in 2026 by using a recursive moving average model, International Journal of Ambient Energy,pp. 1–8, 2022.
2. A. Tomar et al.,Machine learning, advances in computing, renewable energy and communication, 2022.
3. J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, Dataset shift in machine learning.Mit Press, 2022.
4. L.-V. Vlăducu and C.-F. Lăceanu, The impact of renewable energy production on greenhouse gas emissions in romania, Gas,vol. 21, p. 54, 2022.
5. R. Prăvălie, I. S ırodoev, J. Ruiz-Arias, and M. Dumitras¸cu, Using renewable (solar) energy as a sustainable management pathway of lands highly sensitive to degradation in romania. a countrywide analysis based on exploring the geographical and technical solar potentials, Renewable Energy,2022.
6. C. Donnees Mondiales, Roumanie: Données et statistiques du pays, 2022. [Online]. Available: https://www.donneesmondiales.com/europe/roumanie/index.php
7. T. Transelectrica, https://www.transelectrica.ro/ro/web/tel/home,2022.[Online].Available: https://www.transelectrica.ro/ro/web/tel/home
8. A. Ibrahim, R. Kashef, and L. Corrigan, Predicting market movement direction for bitcoin: A comparison of time series modeling methods, Computers & Electrical Engineering,vol. 89, p. 106905, 2021.
9. Z. Cui, J. Wu, Z. Ding, Q. Duan, W. Lian, Y. Yang, and T. Cao, A hybrid rolling grey framework for short time series modelling, Neural Computing and Applications,vol. 33, no. 17, pp. 11 339–11 353, 2021.
10. G. Millán, R. Osorio-Comparán, and G. Lefranc, Preliminaries on the accurate estimation of the hurst exponent using time series, in 2021 IEEE International Conference on Automation/XXIV Congress of the Chilean Association of Automatic Control (ICA-ACCA).IEEE, 2021, pp. 1–8.
11. C. Janiesch, P. Zschech, and K. Heinrich, Machine learning and deep learning, Electronic Markets,vol. 31, no. 3, pp. 685–695, 2021.
12. M. Ramesh, C. Mani, B. Reddy, and M. Venkataramanaiah, Forecasting of bse sensex using simple

exponential smoothing (ses) method, ACA-DEMICIA: An International Multidisciplinary Research Journal,vol. 11, no. 3, pp. 656–665, 2021.
13. M. H. P. Swari, I. P. S. Handika, and I. K. S. Satwika, Comparison of simple moving average, single and modified single exponential smoothing, in 2021 IEEE 7th Information Technology International Seminar (ITIS).IEEE, 2021, pp. 1–5.
14. G. Zhu, L. Li, Y. Zheng, X. Zhang, and H. Zou, Forecasting influenza based on autoregressive moving average and holt-winters exponential smoothing models, Journal of Advanced Computational Intelligence and Intelligent Informatics,vol. 25, no. 1, pp. 138–144, 2021.
15. M. Pleños, Time series forecasting using holt-winters exponential smoothing: Application to abaca fiber data, Zeszyty Naukowe SGGW w Warszawie-Problemy Rolnictwa Swiatowego´ ,vol. 22, no. 2, pp. 17–29, 2022.
16. S. Karlsson, S. Mazur, and H. Nguyen, Vector autoregression models with skewness and heavy tails, arXiv preprint arXiv:2105.11182,2021.
17. F. R. Alharbi and D. Csala, "A seasonal autoregressive integrated moving average with exogenous factors (sarimax) forecasting model-based time series approach, Inventions,vol. 7, no. 4, p. 94, 2022.
18. R. Huang, C. Wei, B. Wang, J. Yang, X. Xu, S. Wu, and S. Huang, Well performance prediction based on long short-term memory (lstm) neural network, Journal of Petroleum Science and Engineering,vol. 208, p. 109686, 2022.
19. J.-Y. Lim, S. Kim, H.-K. Kim, and Y.-K. Kim, "Long short-term memory (lstm)-based wind speed prediction during a typhoon for bridge trafficcontrol, Journal of Wind Engineering and Industrial Aerodynamics,vol. 220, p. 104788, 2022.
20. G. Alotaibi, M. Awawdeh, F. F. Farook, M. Aljohani, R. M. Aldhafiri, and M. Aldhoayan, Artificial intelligence (ai) diagnostic tools: utilizing aconvolutional neural network (cnn) to assess periodontal bone level radiographically—a retrospective study, BMC Oral Health,vol. 22, no. 1, pp. 1–7, 2022.
21. M. Zak, A. Haliuc, S. Cheval, B. Antonescu, A. Tis¸covschi, M. Dobre, F. Tatui, A. Dumitrescu, A. Manea, G. Tudorache et al., Meteorological information at the end of 19 th century from romanian newspapers-an intro to the database," in AGU Fall Meeting Abstracts,vol. 2019, 2019, pp. PP43D–1622.
22. A.-I. Albu, G. Czibula, A. Mihai, I. G. Czibula, S. Burcea, and A. Mezghani, Nextnow: A convolutional deep learning model for the prediction of weather radar data for nowcasting purposes, Remote Sensing,vol. 14, no. 16, p. 3890, 2022.
23. R. Bosneagu, C. E. Lupu, E. Torica, S. Lupu, N. Vatu, V. M. Tanase, C. Vasilache, D. Daneci-Patrau, and I. C. Scurtu, Long-term analysis of air temperatures variability and trends on the romanian black sea coast, Acta Geophysica,vol. 70, no. 5, pp. 2179–2197, 2022.
24. T. Chai and R. R. Draxler, Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature, Geoscientific model development,vol. 7, no. 3, pp. 1247–1250, 2014.
25. M. Elsaraiti and A. Merabet, "Solar power forecasting using deep learning techniques," IEEE Access,vol. 10, pp. 31 692–31 698, 2022.
26. D. Chicco, M. J. Warrens, and G. Jurman, The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation," PeerJ Computer Science,vol. 7, p. e623, 2021.
27. D. S. K. Karunasingha, Root mean square error or mean absolute error? use their ratio as well, Information Sciences,vol. 585, pp. 609–629, 2022.
28. C.-C. Wang, H.-T. Chang, and C.-H. Chien, Hybrid lstm-arma demand-forecasting model based on error compensation for integrated circuit tray manufacturing," Mathematics,vol. 10, no. 13, p. 2158, 2022.
29. A. Rodríguez Sánchez, R. Salmerón Gómez, and C. García, The coefficient of determination in the ridge regression, Communications in statistics-simulation and computation,vol. 51, no. 1, pp. 201–219, 2022.

Todizara Andrianajaina Received the engineer degree in industrial electronics and IT from the Ecole Supérieur Polytechnique Antsiranana, Madagascar, in 2010 and the Advanced study diploma degree for it in the same high school. In December 2017, he obtained his doctorate degree in science and technology, at the University of Antsiranana, Madagascar in co-supervision with the University of Pascal Paolis Corsica, France. Since 2011 to 2017, he was a Research Assistant with the Ecole Supérieur Polytechnique Antsiranana, Madagascar. His research interest includes photovoltaics, electrical engineering, artificial intelligence, machine learning and forecasting. Dr. T. Andrianajaina is actively involved in development projects with his university or with international partners. This year he coordinated a project to strengthen the capacity of the doctoral schools of the University of Antsiranana.

Tsivalalaina David Razafimahefa received his PhD in Automation, Signal, Generation and Robotics from University of Corsica Pasquale Paoli, France, and his PhD in Electrical Engineering from the University of Antsiranana, Madagascar, in 2016. Since 2017, he has been alecturer at Antsiranana University, teaching electrical and technical computer science. His research project is focused on prognosis and health management (PHM) of electrical systems.

Cristian-Gyozo Haba Received his Ph.D. in Automatic Control in 2000 and Engineer Degree in Electrical Engineering in 1988 from “Gheorghe Asachi” Technical University of Iasi (Romania). Since 1990 he is teaching at Faculty of Electrical Engineering, “Gheorghe Asachi” Technical University of Iasi where he is now a Professor in the Department of Electrical Engineering. At present he works on EDA, micro-controller and FPGA based embedded systems and e-learning. He is an IEEE, ACM and EAI member involved in several research projects on embedded system design, distributed measurement systems, home building automation, ambient assisted living and IoT.

Dorin Dumitru Lucache Dorin-Dumitru Lucache received the M.S. and Ph.D. degrees from the ” Gheorghe Asachi”, Technical University of Iasi, Romania, in 1986 and 2001, respectively, both in electric engineering. He received also the M.S. in Mathematics and Business Administration from the ” Alexandru Ioan Cuza” University of Iasi, Romania, in 1994 and 2007, respectively. Full professor on Utilizations of Electrical Energy from 2009. He is a Member of the Institute of Electrical and Electronics Engineers (IEEE #85012847) and AGIR (General Association of Engineers of Romania) and organizer of several international conferences in the field of electrical and power engineering held in Iasi, Romania and Chisinau, Republic of Moldova. The research interest has spiraled through various aspects of the electrical engineering, from electric lighting, electric heating and cooling, energetic efficiency, optimal design of electromagnetic devices, renewable energy sources, smart grids and up to devices for diagnostic based on alternative and complementary medicine or functional electric stimulation.