Regression-based method for real-time solar power plant efficiency forecasting

Myroslav Komar[1,*,†], Khrystyna Lipianina-Honcharenko[1,†], Valentyn Domanskyi[1,†] and Nazar Melnyk[1,†]

[1] *West Ukrainian National University, Lvivska str., 11, Ternopil, 46009, Ukraine*

Abstract

The importance of this research lies in the growing reliance on solar energy as a key renewable energy source. Solar power plants offer low operational costs, ease of maintenance, and substantial reliability, making them an attractive option for clean energy production. However, the efficiency of these plants can be significantly influenced by external factors such as weather conditions and the physical characteristics of the solar panels. The paper elaborates on various forecasting horizons-ranging from very short-term to long-term-and discusses the suitability of different models like artificial neural networks, time-series forecasting, machine learning, and ensemble methods for these applications. Utilizing data from a solar power plant the study tests several regression models to identify the one with the best forecasting accuracy. The Gradient Boosting Regressor emerged as the most effective model, demonstrating its potential in accurately predicting solar power output. The methodology's success highlights the possibility of integrating solar power plants more efficiently into smart grid systems and optimizing energy management practices. The paper presents a robust method for real-time forecasting of solar power plant efficiency that could significantly benefit energy management and the integration of renewable energy sources into power systems. It opens avenues for further research into improving forecasting techniques and underscores the critical role of accurate prediction models in the advancement of renewable energy technologies.

Keywords

Solar energy, forecasting, regression model, solar panel efficiency 1

1. Introduction

With the growing popularity of clean energy, solar power plants have become one of the leading renewable sources [14]. Solar power plants have a fairly low cost, are easy to operate, and have high reliability and durability. This and a number of other economic factors have led to a significant increase in the number of households and communities in Ukraine equipped with solar power plants.

However, due to a large number of external factors, the efficiency of solar power plants is quite unpredictable. Weather factors (solar radiation, air temperature, cloud cover, humidity) play a key role in their operation [13, 19]. The position of the solar panels themselves, their type and characteristics also have an impact. Therefore, an important task is to introduce systems for forecasting the generated electricity for more efficient energy management, as well as to simplify the process of integrating solar power plants into Smart Grid systems [5, 6, 21].

Smart Grid systems are becoming increasingly relevant in the world of modern energy technologies, as they play a key role in transitioning to a sustainable and efficient energy infrastructure. Let's consider several aspects that highlight the importance and relevance of Smart Grid:

1. Integration of renewable energy sources: Smart grids facilitate the more efficient integration of various renewable sources, such as solar panels and wind turbines, enabling quick response to changes in energy production, which is inherently variable in nature.
2. Enhancing energy system reliability: Smart Grids ensure greater reliability of the energy system by early detection and automatic rectification of faults in the power grid. This helps minimize downtime and impacts on consumers.
3. Efficiency in energy flow management: With the aid of modern data analysis and management technologies, Smart Grids allow for the optimization of energy distribution according to demand, time of day, and other factors, thereby reducing costs and energy consumption.
4. Improvement of consumer experience: Smart meters and consumption management systems integrated into Smart Grids provide consumers with detailed information about their energy consumption, enabling them to better plan their expenses and usage.
5. Climate change adaptation: Smart Grids play a crucial role in combating climate change by optimizing the use of renewable sources and reducing dependence on fossil fuels, thus contributing to the reduction of greenhouse gas emissions.
6. Energy security: Smart Grids help ensure energy security at the national level by enabling efficient responses to energy crises and rapidly adapting resources in the face of energy challenges or instability. With analytical capabilities and real-time data, Smart Grids can redistribute energy within the grid, reducing the risk of major outages or supply disruptions.
7. Engaging consumers in energy system management: Smart Grids empower consumers to become active participants in the energy market through home solar installations, energy storage systems, or participation in demand response programs. This not only improves energy efficiency but also promotes the decentralization of energy resources.

Given these advantages, Smart Grids are a necessary condition for modern energy systems seeking stability, economic efficiency, and reduced environmental impact. These systems play a crucial role in shaping the future energy landscape and addressing global energy challenges.

2. State of the art

Currently, there are a large number of models and methods for predicting the efficiency of solar power plants [4]. Depending on the duration of forecasts, forecasting can be divided into:

1. Very short term forecasting. This horizon focuses on immediate future predictions, ranging from seconds to less than 30 minutes. It's crucial for real-time energy management, enabling operators to respond swiftly to sudden changes in solar power generation. This forecasting is pivotal for maintaining grid stability, especially in systems with significant solar penetration, by aiding in the instantaneous balancing of supply and demand. Very short-term forecasts are utilized for dynamic grid operations, including real-time electricity dispatch, power smoothing, and PV storage control. They assist in optimizing the grid's responsiveness to fluctuations in solar energy production, ensuring efficient and reliable energy delivery.

2. Short-Term forecasting. Short-term forecasting typically spans from 30 minutes to several hours or days. It's essential for day-to-day operational planning and energy market transactions. Accurate short-term forecasts help in scheduling power plant operations, managing energy storage systems, and facilitating efficient energy trading. Applications: This horizon supports economic load dispatch, power system operation, and the integration of renewable energies into power management systems. It helps utilities and grid operators plan for energy production, distribution, and consumption, minimizing operational costs and enhancing grid reliability [8].
3. Intra-Day Forecasting. Intra-day forecasting, covering a span of 1 to 6 hours, bridges the gap between short-term and medium-term horizons. It's particularly useful for managing energy supply and demand within the same day, offering insights into how solar power output will vary over several hours. This forecasting horizon is valuable for electricity trading, where energy prices fluctuate throughout the day, and for managing zone-specific electric loads. It aids in making informed decisions on energy procurement, storage, and distribution within daily operational cycles.
4. Medium-Term Forecasting. Medium-term forecasting ranges from 6 to 24 hours, extending up to a week or month in some contexts. It provides a broader outlook on solar power generation, enabling strategic decisions related to maintenance scheduling, energy procurement, and system optimization. Utilities and energy managers use medium-term forecasts for maintenance planning of power systems, ensuring optimal operation of equipment and integration of solar power. It also supports better planning for energy trading and load forecasting over a week or month.
5. Long-Term Forecasting. Long-term forecasts predict solar power generation beyond 24 hours, often extending to months or a year. This horizon is critical for strategic planning, investment decisions, and policy formulation, offering a long-range view of energy generation trends and potential impacts on the grid. Long-term forecasting is used for capacity planning, investment analysis, and policy-making. It helps stakeholders anticipate future energy production, enabling informed decisions on infrastructure development, resource allocation, and renewable energy integration strategies.

Depending on the required forecasting horizon and the available data, the required forecasting method will depend on the forecasting method. The methods can be divided into the following:

• Artificial neural networks. ANNs are computational models inspired by the human brain's network of neurons. They consist of input, hidden, and output layers with interconnected nodes (neurons) that process data and can learn from it. ANNs are particularly effective in handling nonlinear and complex data patterns, making them ideal for forecasting tasks where traditional linear models fall short [3]. Typically used for short-term forecasting, ANNs can predict solar power output for a few hours to several days ahead, adapting to the dynamic nature of solar irradiance.
• Time-series forecasting methods. This approach includes methods such as exponential smoothing, autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA). Time series analysis aims to predict future values by analyzing patterns in past data [8]. It was found that the ARIMA model more accurately predicts the operation of a solar station for a 24-hour period compared to other models.
• Machine learning. These methods rely on the ability of AI to learn from historical data and improve predictions through iterative learning [2]. The main methods in this category include multilayer perceptron neural networks (MLPNN), recurrent neural networks

(RNN), and feed-forward neural networks. Machine learning is used for forecasting in different ranges - from 30 minutes to 24 hours.

- Physical and statistical models. These methods, based on solar panel characteristics and historical data, use mathematical equations to extract patterns and correlations to reduce error and increase forecasting accuracy. These methods include curve fitting, moving average (MA), and autoregressive models (AR). Usually, these methods are used for forecasting from one day. The forecast can be extended for several months or years [4, 9].
- Ensemble methods combine multiple individual models to improve forecast accuracy and reliability. By leveraging the strengths and mitigating the weaknesses of each model, ensemble techniques often achieve better performance than any single forecasting method. This approach can integrate diverse models like ANNs, time-series methods, and other machine learning algorithms. While commonly applied to medium-term forecasting, ensemble techniques can be tailored for short-term and long-term forecasting, offering a versatile solution for various prediction needs.
- Nowcasting (Intra-Hour Forecasting). Nowcasting focuses on predicting the immediate future, utilizing real-time data to make short-term predictions. This approach is crucial for applications where timely and accurate forecasts are essential, such as in managing solar power grids where sudden changes in solar irradiance can impact grid stability. Very short-term, concentrating on the imminent future from seconds to up to an hour, making it indispensable for operational decision-making in energy management and grid control.

3. Statement of the Research Problem

The primary objective of this research is to develop a robust method for real-time forecasting of solar power plant efficiency utilizing regression models. Given the unpredictable efficiency of solar power plants due to external factors like weather conditions and the physical characteristics of solar panels, it is crucial to enhance forecasting accuracy to support efficient energy management and integration into smart grid systems. The task involves testing various regression models, including Gradient Boosting Regressor, to determine which model provides the best accuracy in predicting solar power output. This research aims to provide insights into optimizing solar power plant operations and advancing renewable energy technologies.

The subject of this research is the development and analysis of regression models for forecasting the efficiency of solar power plants. The study focuses on the impact of external factors, such as weather conditions and the state of solar panels, on the accuracy of predicting solar energy production. Particular attention is given to the adaptation and optimization of regression models, especially the Gradient Boosting Regressor, for real-time use within the context of energy management and integration of solar power plants into smart grid systems.

The object of this research is solar power plants and their operation under various weather and operational conditions. The study explores the specifics and dynamics of the efficiency of solar panels, which are influenced by external factors such as solar radiation and temperature changes. The impact of these factors on energy production is analyzed, which is critically important for developing more accurate forecasting methods and optimizing the operation of solar power plants in the context of modernizing energy infrastructure.

This research makes a significant contribution to the field of solar power plant efficiency forecasting by employing innovative regression models optimized for real-time operation. The core scientific novelty lies in the development of a methodology that integrates external factors, such as weather conditions and the condition of solar panels, directly into the forecasting model. Utilizing the Gradient Boosting method along with other regression models tailored to minimize forecasting errors, the study addresses key limitations of standard approaches that often fail to

consider the complexity and dynamics of solar power plant operational characteristics. This approach can significantly enhance the accuracy of energy flow management in conditions of unstable energy production, which is crucial for the integration of solar power plants into smart grid systems and improving the overall efficiency of renewable energy utilization.

4. Dataset

The data on electricity generation and station temperature were obtained from a solar power plant located in Zelene village, Husiatyn district, Ternopil region (lat: 49.313965, long: 26.098843). The station is shown in Figure 1. The rated capacity of the plant is 30kW. The generated energy is converted using 3 inverters. The output power data of the power plant is collected for each inverter separately.

Figure 1: Photo of the solar power plant in Zelene village

The forecasting uses data from a solar power plant obtained from January 1, 2019 to December 31, 2022. The granularity of the data is 30 seconds. The data is available only for the period of the day when there was power at the inverters' output (practically equal to the duration of daylight hours).

4.1. Factors affecting the forecast of solar power plant efficiency

The Open Meteo website [10] was used to obtain historical weather indicators. Based on the literature analysis, a list of weather indicators that have a significant impact on the quality of electricity generation forecasts was identified:

- air temperature (°C)
- humidity (%)
- atmospheric pressure(hPa)
- wind speed at a height of 10 meters above the earth's surface (km/h)
- percentage of cloudiness (%)
- percentage of cloud transparency (%)
- direct solar radiation (W/m^2)
- scattered solar radiation (W/m^2)

The granularity of the available data is 1 hour. To obtain historical data on the position of the sun in the specified coordinates (zenith, azimuth, and altitude), we used the Solcast.com website [17]. The following indicators were also used from this resource:

- air temperature (°C)
- air humidity (%)
- atmospheric pressure (hPa)
- radiation values obtained by the Clear Sky method (W/m^2) [16]

The granularity of the available data is 30 seconds. To check the existing dependence between the selected indicators and the efficiency of a solar power plant, a correlation matrix was built based on the Pearson correlation coefficient (Figure 2).
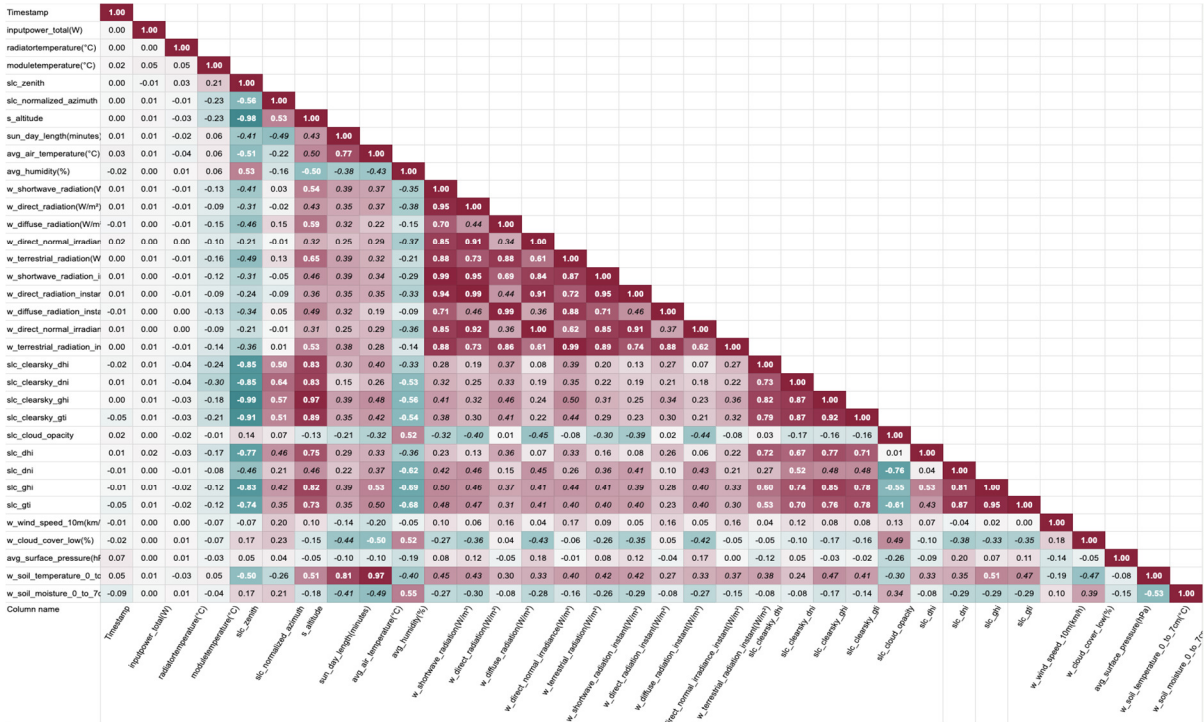
	Timestamp	inputpower_total(W)	radiatortemperature(°C)	moduletemperature(°C)	slc_zenith	slc_normalized_azimuth	s_altitude	sun_day_length(minutes)	avg_air_temperature(°C)	avg_humidity(%)	w_shortwave_radiation(W)	w_direct_radiation(W/m²)	w_diffuse_radiation(W/m)	w_direct_normal_irradian	w_terrestrial_radiation(W)	w_shortwave_radiation_i	w_direct_radiation_instan	w_diffuse_radiation_insta	w_direct_normal_irradian_in	w_terrestrial_radiation_in	slc_clearsky_dhi	slc_clearsky_dni	slc_clearsky_ghi	slc_clearsky_gti	slc_cloud_opacity	slc_dhi	slc_dni	slc_ghi	slc_gti	w_wind_speed_10m	w_cloud_cover_low(%)	avg_surface_pressure(hPa)	w_soil_temperature_0_to	w_soil_moisture_0_to_7c
Timestamp	1.00																																	
inputpower_total(W)	0.00	1.00																																
radiatortemperature(°C)	0.00	0.00	1.00																															
moduletemperature(°C)	0.02	0.05	0.05	1.00																														
slc_zenith	0.00	-0.01	0.03	0.21	1.00																													
slc_normalized_azimuth	0.00	0.01	-0.01	-0.23	-0.56	1.00																												
s_altitude	0.00	0.01	-0.03	-0.23	-0.98	0.53	1.00																											
sun_day_length(minutes)	0.01	0.01	-0.02	0.06	-0.41	-0.49	0.43	1.00																										
avg_air_temperature(°C)	0.03	0.01	-0.04	0.06	-0.51	-0.22	0.50	0.77	1.00																									
avg_humidity(%)	-0.02	0.00	0.01	0.06	0.53	-0.16	-0.50	-0.38	-0.43	1.00																								
w_shortwave_radiation(W)	0.01	0.01	-0.01	-0.13	-0.41	0.03	0.54	0.39	0.37	-0.35	1.00																							
w_direct_radiation(W/m²)	0.01	0.01	-0.01	-0.09	-0.31	-0.02	0.43	0.35	0.37	-0.38	0.95	1.00																						
w_diffuse_radiation(W/m)	-0.01	0.00	-0.01	-0.15	-0.46	0.15	0.59	0.32	0.22	-0.15	0.70	0.44	1.00																					
w_direct_normal_irradian	0.02	0.00	0.00	-0.10	-0.21	-0.01	0.32	0.25	0.29	-0.37	0.85	0.91	0.34	1.00																				
w_terrestrial_radiation(W)	0.00	-0.01	-0.01	-0.16	-0.49	0.13	0.65	0.39	0.32	-0.21	0.88	0.73	0.88	0.61	1.00																			
w_shortwave_radiation_i	0.01	0.00	-0.01	-0.12	-0.31	-0.05	0.46	0.39	0.34	-0.29	0.99	0.95	0.69	0.84	0.87	1.00																		
w_direct_radiation_instan	0.01	0.00	-0.01	-0.09	-0.24	-0.09	0.36	0.35	0.35	-0.33	0.94	0.99	0.44	0.91	0.72	0.95	1.00																	
w_diffuse_radiation_insta	-0.01	0.00	0.00	-0.13	-0.34	0.05	0.49	0.32	0.19	-0.09	0.71	0.46	0.88	0.36	0.71	0.46	0.37	1.00																
w_direct_normal_irradian_in	0.01	0.00	0.00	-0.09	-0.21	-0.01	0.31	0.25	0.29	-0.36	0.85	0.92	0.36	1.00	0.62	0.85	0.91	0.37	1.00															
w_terrestrial_radiation_in	0.00	-0.01	-0.01	-0.14	-0.36	0.01	0.53	0.38	0.28	-0.14	0.88	0.73	0.86	0.61	0.99	0.89	0.74	0.88	0.62	1.00														
slc_clearsky_dhi	-0.02	0.01	-0.04	-0.24	-0.85	0.50	0.83	0.30	0.40	-0.33	0.28	0.19	0.37	0.08	0.39	0.20	0.13	0.27	0.07	0.27	1.00													
slc_clearsky_dni	0.01	0.01	-0.04	-0.30	-0.85	0.64	0.83	0.15	0.26	-0.53	0.32	0.25	0.33	0.19	0.35	0.22	0.19	0.21	0.18	0.22	0.73	1.00												
slc_clearsky_ghi	-0.01	0.01	-0.03	-0.18	-0.99	0.57	0.97	0.39	0.48	-0.56	0.41	0.32	0.46	0.24	0.50	0.31	0.25	0.34	0.23	0.36	0.82	0.87	1.00											
slc_clearsky_gti	-0.05	0.01	-0.03	-0.21	-0.91	0.51	0.89	0.35	0.42	-0.54	0.38	0.30	0.41	0.22	0.44	0.29	0.23	0.30	0.21	0.32	0.79	0.87	0.92	1.00										
slc_cloud_opacity	0.02	0.00	-0.02	-0.01	0.14	0.07	-0.13	-0.21	-0.32	0.52	-0.32	-0.40	0.01	-0.45	-0.08	-0.30	-0.39	0.02	-0.44	-0.08	0.03	-0.17	-0.16	-0.16	1.00									
slc_dhi	-0.01	0.02	-0.03	-0.17	-0.77	0.46	0.75	0.29	0.33	-0.36	0.23	0.13	0.36	0.07	0.33	0.16	0.08	0.26	0.06	0.22	0.72	0.67	0.77	0.71	0.01	1.00								
slc_dni	-0.01	0.00	-0.01	-0.08	-0.46	0.21	0.46	0.22	0.37	-0.62	0.42	0.46	0.15	0.45	0.26	0.36	0.41	0.10	0.43	0.21	0.27	0.52	0.48	0.48	-0.76	0.04	1.00							
slc_ghi	-0.01	0.01	-0.02	-0.12	-0.83	0.42	0.82	0.39	0.53	-0.69	0.50	0.46	0.37	0.41	0.44	0.41	0.39	0.28	0.40	0.33	0.60	0.74	0.85	0.78	-0.55	0.53	0.81	1.00						
slc_gti	-0.05	0.01	-0.02	-0.12	-0.74	0.35	0.73	0.35	0.50	-0.68	0.48	0.47	0.31	0.41	0.40	0.40	0.40	0.23	0.40	0.30	0.53	0.70	0.76	0.78	-0.51	0.43	0.87	0.95	1.00					
w_wind_speed_10m(km)	-0.01	0.00	0.00	-0.07	-0.07	0.20	0.10	-0.14	-0.20	-0.05	0.10	0.06	0.16	0.04	0.17	0.09	0.05	0.16	0.05	0.16	0.04	0.12	0.08	0.08	0.13	0.07	-0.04	0.02	0.00	1.00				
w_cloud_cover_low(%)	-0.02	0.00	0.01	0.07	0.17	0.23	-0.15	-0.44	-0.50	0.52	-0.27	-0.36	0.04	-0.43	-0.06	-0.26	-0.35	0.05	-0.42	-0.05	-0.05	-0.10	-0.17	-0.16	0.49	-0.10	-0.38	-0.33	-0.35	0.18	1.00			
avg_surface_pressure(hPa)	0.07	0.00	0.01	-0.03	0.05	0.04	-0.05	-0.10	-0.10	-0.19	0.08	0.12	-0.05	0.18	-0.01	0.08	0.12	-0.04	0.17	0.00	-0.12	0.05	-0.03	-0.02	-0.26	-0.09	0.20	0.07	0.11	-0.14	-0.05	1.00		
w_soil_temperature_0_to	0.05	0.01	-0.03	0.05	-0.50	-0.26	0.51	0.81	0.97	-0.40	0.45	0.43	0.30	0.33	0.40	0.42	0.42	0.27	0.33	0.37	0.38	0.24	0.47	0.41	-0.30	0.33	0.35	0.51	0.47	-0.19	-0.47	-0.08	1.00	
w_soil_moisture_0_to_7c	-0.09	0.00	-0.04	0.17	0.21	-0.18	-0.41	-0.49	0.55	-0.27	-0.30	-0.08	-0.28	-0.16	-0.26	-0.29	-0.08	-0.27	-0.15	-0.08	-0.08	-0.17	-0.14	0.34	-0.08	-0.29	-0.29	0.10	0.39	-0.15	0.39	-0.15	-0.53	1.00

Figure 2: Correlation matrix based on Pearson's correlation coefficient

4.2. Dataset normalization

Since the data have different granularity and duplicate indicators, the following steps were taken to normalize them:

1. The total capacity of the solar power plant was calculated by summing the removal of individual inverters.
2. The sun's azimuth is converted to a modulo value. Thus, a scale from 0 to 180 points was obtained, which corresponds to the deviation of the sun from the south (Figure 3).
3. The average air temperature, humidity, and atmospheric pressure are calculated between Open Meteo and Solcast.
4. Using the SunCalc library [18], we calculated the duration of daylight hours in minutes.
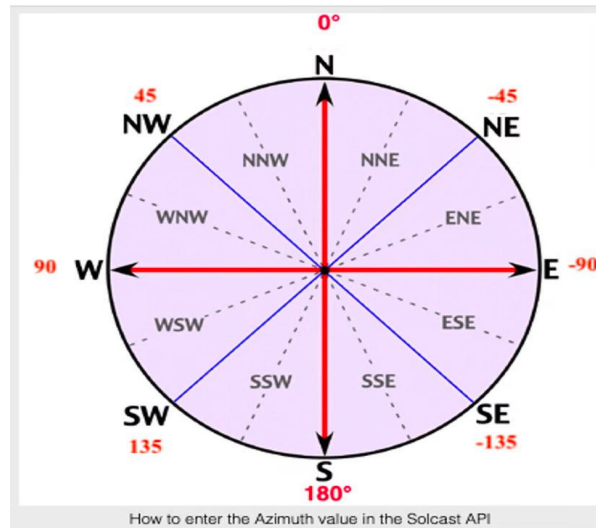5. All indicators are reduced to a granularity of one hour by calculating the average value.

Figure 3: Scale of sun deviation from the south [15]

As a result, we obtained a data set of 18840 records, which consists of the following data:

- power of the solar station (W)
- radiator temperature (°C)
- module temperature (°C)
- sunrise
- sun azimuth
- height of the sun above sea level
- air temperature (°C)
- air humidity (%)
- atmospheric pressure (hPa)
- wind speed (km/h)
- percentage of cloud cover (%)
- percentage of cloud transparency (%)
- direct solar radiation (W/m^2)
- scattered solar radiation (W/m^2)
- radiation values obtained by the Clear Sky method (W/m^2)

The data granularity is 1 hour.

5. Regression Model for Predicting the Efficiency of a Solar Power Plant

One of the research objectives is to predict online performance, which is adjusted by updating weather indicators. Therefore, it was decided to use regression models for forecasting, which are faster than neural networks and their input data can be variable.

To examine the regression models results quantity, Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) were used.

MSE measures the average of the squares of the errors – that is, the average squared difference between the estimated values and the actual value. MSE is calculated as the mean of the squared differences between predicted and actual values. MSE is very sensitive to large errors due to squaring each term, which means it gives a relatively high weight to large errors. This property can be very useful when large errors are particularly undesirable.

MAE measures the average magnitude of the errors in a set of predictions, without considering their direction (i.e., it takes the average over the absolute values of the errors). MAE is calculated as the mean of the absolute differences between predicted and actual values. MAE is particularly

robust to outliers as it does not square the errors before summing them. It provides a linear score that reflects the average error magnitude.

MAPE expresses the accuracy as a percentage, and it measures the average absolute percent error for each prediction error compared to the actual value. MAPE is calculated as the mean of the absolute differences between the predicted and actual values, divided by the actual values, typically expressed as a percentage. MAPE is easy to interpret as a percentage, which makes it straightforward for communicating model performance.

The forecasting took place in several stages.

1. The interquartile range (IQR) method was used to identify and remove outliers.
2. The data was divided into training (January 1, 2019 - December 31, 2021) and test (January 1, 2022 - December 31, 2022) networks. Figure 4 shows the distribution of data by the training library.
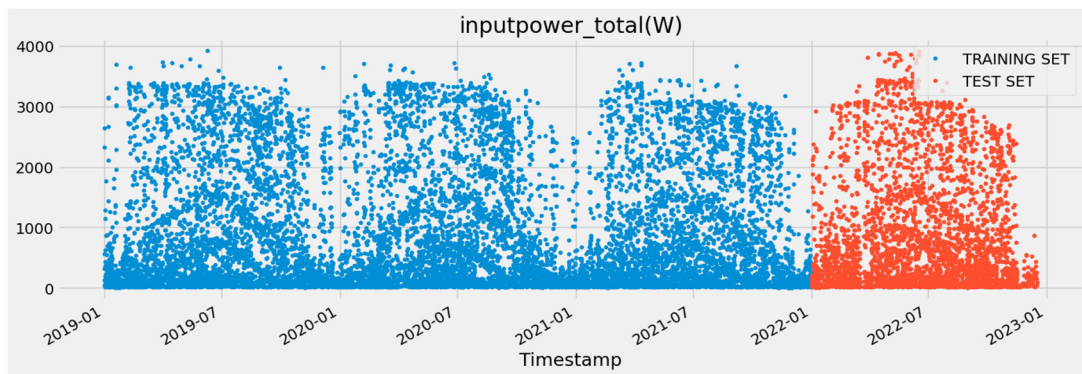
Figure 4: Data distribution by the educational library

3. Additional time identifiers (hour, day, quarter) were created for the test dataset.
4. Several regression models were used for forecasting. Each model is trained on a training dataset and makes predictions on a test dataset. Table 1 compares the quality metrics of the models used.
5. The GridSearchCV function was used to select the model parameters with the best performance.
6. Based on the best of the selected Gradient Boosting Regressor models, the forecasting was performed. The quality indicators were:

Mean Squared Error: 488590.0169641469
Mean Absolute Error: 526.9728465745633
Mean Absolute Percentage Error: 184.08511806925392

Table 1
Comparison of quality metrics of the models used

Model	MSE	MAE	MAPE
LightGBM	568934.1545839079	528.9011747899133	154.2357938049348
XBGRegressor	650515.7686498845	567.5401797341103	186.9378575677895
CatBoost	604615.6297283806	564.6730864737467	200.8061044696372
Random Forest	704809.2438455686	556.3300289324546	138.8617765424779
Gradient Boosting	497419.2943262356	519.460641187808	164.4226289699118

AdaBoost	585418.3816034624	616.2476624437463	293.4843335358631
SVM	1129798.3849226362	728.1819757662427	310.78685738411315

Figure 5 shows a comparison of real and predicted data. Forecasting errors were calculated and analyzed by day.
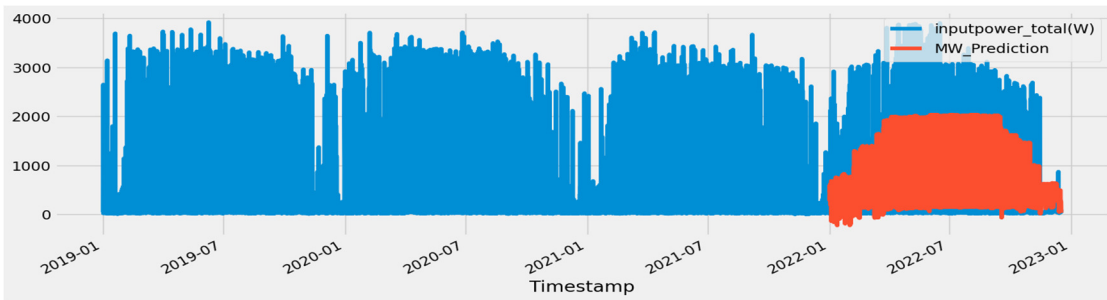

Figure 5: Comparison of real and predicted data

7. Future time stamps and time features for the future period were generated, based on which the forecasting was performed. The forecasting results are shown in Figure 6.
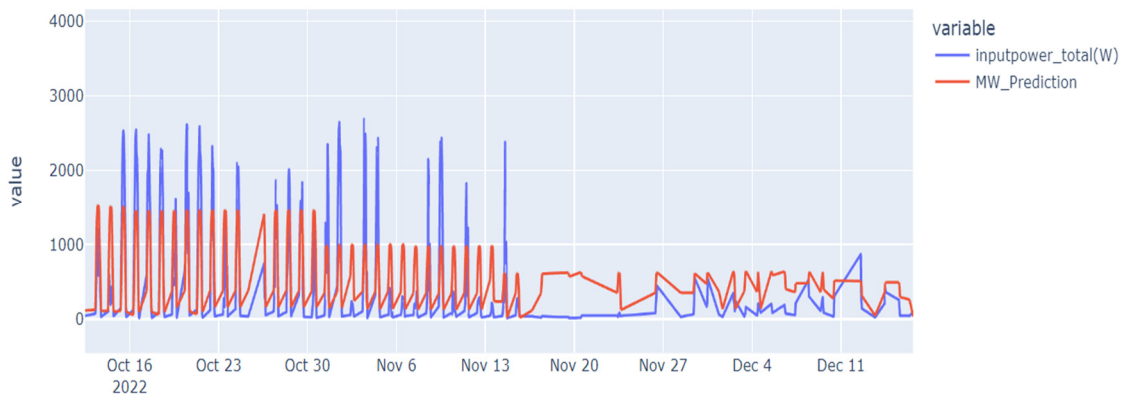
Figure 6: Forecasting results

Using the Gradient Boosting Regressor regression model, we obtained fairly accurate forecasting results. Larger deviations in winter are explained by the shorter daylight hours when solar panels are in operation. Accordingly, this leads to a reduction in the amount of data on which to base forecasts. This is also affected by stable weather and the presence of snow on the panels, which reduces their efficiency. Other researchers have also noted similar deviations in their work [1, 7, 12].

Conclusions

This study presents a novel method for real-time forecasting of solar power plant efficiency, leveraging regression models to predict performance based on historical weather data and the solar power plant's operational data. By examining the forecasting accuracy of various models, the research identifies the Gradient Boosting Regressor as the most effective, enabling more efficient integration of solar power into smart grid systems and optimizing energy management practices.

The research emphasizes the importance of accurate and real-time efficiency predictions for solar power plants, which face efficiency variability due to external factors like weather conditions and the physical characteristics of solar panels. Through comprehensive data

normalization and testing of several regression models, the study offers insights into the critical role of weather indicators and solar panel positioning in solar power generation. It highlights the potential of regression models, particularly the Gradient Boosting Regressor, in enhancing forecasting accuracy, thereby supporting better energy management and integration of renewable energy sources into power systems.

The method developed in this study is a significant contribution to the field of renewable energy, providing a robust framework for predicting the efficiency of solar power plants in real-time. It opens up new possibilities for research into improving forecasting techniques and underscores the vital importance of accurate prediction models in advancing renewable energy technologies. Future efforts should focus on addressing the challenges of seasonal variations and external factors like snow coverage to further refine predictive capabilities.

References

[1] F.-V. Gutiérrez-Corea, M. Á. M. Callejo, M.-P. Moreno-Regidor, and M.-T. Manrique-Sancho. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. Solar Energy, vol. 134, pp. 119–131, Sep. 2016, doi: 10.1016/j.solener.2016.04.020.

[2] Golovko, V., Kroshchanka, A., Bezobrazov, S., Komar, M., Novosad, O. Development of Solar Panels Detector. 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 - Proceedings, 2018, pp. 761–764, 8632132, doi: 10.1109/INFOCOMMST.2018.8632132

[3] Golovko, V., Kroshchanka, A., Mikhno, E., Komar, M., Sachenko, A. Deep convolutional neural network for detection of solar panels. Lecture Notes on Data Engineering and Communications Technologies, 2021, 48, pp. 371–389, doi: 10.1007/978-3-030-43070-2_17

[4] H. Zang et al. Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. Iet Generation Transmission & Distribution, vol. 12, no. 20, pp. 4557–4567, Sep. 2018, doi: 10.1049/iet-gtd.2018.5847.

[5] I. Colak, "Introduction to smart grid,"2016 International Smart Grid Workshop and Certificate Program (ISGWCP), Istanbul, Turkey, 2016, pp. 1-5, doi: 10.1109/ISGWCP.2016.7548265.

[6] IEEE Smart Grid: The Leaders in Smart Grid Technology. URL: https://smartgrid.ieee.org

[7] M. Ding, L. Wang, and R. Bi. An ANN-based approach for forecasting the power output of photovoltaic system. Procedia Environmental Sciences, vol. 11, pp. 1308–1315, Jan. 2011, doi: 10.1016/j.proenv.2011.12.196.

[8] M. Elsisi, M. Amer, A. Dababat, and C. Su. A comprehensive review of machine learning and IoT solutions for demand side energy management, conservation, and resilient operation. Energy, vol. 281, p. 128256, Oct. 2023, doi: 10.1016/j.energy.2023.128256.

[9] M. G. De Giorgi, P. M. Congedo, and M. Malvoni. Photovoltaic power forecasting using statistical methods: impact of weather data. Iet Science Measurement & Technology, vol. 8, no. 3, pp. 90–97, May 2014, doi: 10.1049/iet-smt.2013.0135.

[10] Open Meteo. URL: https://open-meteo.com/.

[11] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renewable & Sustainable Energy Reviews, vol. 124, p. 109792, May 2020, doi: 10.1016/j.rser.2020.109792.

[12] S. Al-Dahidi, O. Ayadi, M. Alrbai, and J. Adeeb. Ensemble Approach of Optimized Artificial Neural networks for Solar Photovoltaic power prediction. IEEE Access, vol. 7, pp. 81741–81758, Jan. 2019, doi: 10.1109/access.2019.2923905.

[13] S. Ghazi and K. Ip. The effect of weather conditions on the efficiency of PV panels in the southeast of UK. Renewable Energy, vol. 69, pp. 50–59, Sep. 2014, doi: 10.1016/j.renene.2014.03.018.

[14] S. Pfenninger and I. Staffell. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy, vol. 114, pp. 1251–1265, Nov. 2016, doi: 10.1016/j.energy.2016.08.060.

[15] Solcact Azimuth. URL: https://kb.solcast.com.au/azimuth.

[16] Solcast irradiance and weather methodology. URL: https://solcast.com/irradiance-data-methodology.

[17] Solcast. URL: https://toolkit.solcast.com.au.

[18] SunCalc. URL: https://www.npmjs.com/package/suncalc.

[19] T. Ishii, K. Otani, T. Takashima, and X. Yang. Solar spectral influence on the performance of photovoltaic (PV) modules under fine weather and cloudy weather conditions. Progress in Photovoltaics: Research and Applications, vol. 21, no. 4, pp. 481–489, Nov. 2011, doi: 10.1002/pip.1210.

[20] V. Krylov et al. Multiple Regression Method for Analyzing the Tourist Demand Considering the Influence Factors. 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Metz, France, 2019, pp. 974-979, doi: 10.1109/IDAACS.2019.8924461.

[21] Y. Wang, Q. Chen, T. Hong and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," in IEEE Transactions on Smart Grid, vol. 10, no. 3, pp. 3125-3148, May 2019, doi: 10.1109/TSG.2018.2818167.