Süleyman Demirel Üniversitesi
YEKARUM e-DERGİ
(Journal of YEKARUM)
Cilt 9, Sayı 2, 128-146, 2024
E - ISNN:1309-9388

Machine Learning-Based Energy Forecasting for PV Power Plants

Muhammed TAMAY[1], Gul Fatma TURKER[2*]

[1] Suleyman Demirel University, Graduate School of Natural and Applied Sciences, Department of Computer Engineering, 3200, Isparta, Türkiye, (ORCID: 0009-0006-5872-5866), mhmmdtamay@gmail.com
[2*] Suleyman Demirel University, Faculty of Engineering, Department of Computer Engineering, 3200, Isparta, Türkiye (ORCID: 0000-0001-5714-5102), gulturker@sdu.edu.tr

(İlk Geliş Tarihi 23/11/2024 ve Kabul Tarihi 18/12/2024)

ABSTRACT

As the global population continues to grow and technological advancements progress, energy demand is becoming increasingly noticeable worldwide. Solar energy, which is among the sustainable energy sources to reduce the environmental impact of fossil fuels, has a critical role in the global energy transition. Energy generation forecasts play a vital role in supply-demand balance, grid stability and cost optimization. Moreover, accurate and reliable generation forecasts are essential to facilitate the integration of renewable energy sources and improve the efficiency of energy systems. In this study, generation data from April 2022 to April 2024 for a solar power plant in Denizli province and weather data obtained from Solcast API service are used. The performance of machine learning algorithms such as XGBoost, Extra Trees, k-Nearest Neighbors (KNN), Gradient Boosting, Random Forest and Linear Regression are evaluated. The results show that the KNN model outperforms the other algorithms with a Mean Square Error (MSE) of 112.282, Root Mean Square Error (RMSE) of 10.596, Mean Absolute Error (MAE) of 4.094 and R^2 score of 0.969. This study contributes to a more reliable estimation of solar power generation, facilitating the integration of renewable energy sources and offering significant potential for the optimization of energy management systems.

Keywords: *PV Plants, Solar Power Generation Forecasting, Machine Learning, Renewable Energy*

PV Enerji Santralleri için Makine Öğrenmesi Tabanlı Enerji Üretim Tahmini

ÖZET:

Dünya nüfusu ve teknolojik gelişmelerin sürekli artmasıyla birlikte, enerji talebi tüm dünyada giderek daha fazla hissedilmektedir. Fosil yakıtların çevresel etkilerini azaltmak amacıyla sürdürülebilir enerji kaynakları arasında yer alan güneş enerjisi, küresel enerji dönüşümünde kritik bir öneme sahiptir. Enerji üretim tahminleri, arz-talep dengesi, şebeke stabilitesi ve maliyet optimizasyonu açısından hayati bir rol oynamaktadır. Ayrıca, doğru ve güvenilir üretim tahminleri, yenilenebilir enerji kaynaklarının entegrasyonunu kolaylaştırmak ve enerji

* Corresponding author E-mail: gulturker@sdu.edu.tr

sistemlerinin verimliliğini artırmak için gereklidir. Bu çalışmada, Denizli ilindeki bir güneş enerji santraline ait Nisan 2022 ile Nisan 2024 arasındaki üretim verileri ve Solcast API servisinden elde edilen hava durumu verileri kullanılmıştır. XGBoost, Extra Trees, k-Nearest Neighbors (KNN), Gradient Boosting, Random Forest ve Linear Regression gibi makine öğrenmesi algoritmalarının performansı değerlendirilmiştir. Elde edilen sonuçlar, KNN modelinin Ortalama Kare Hata (MSE) değeri 112.282, Kök Ortalama Kare Hata (RMSE) değeri 10.596, Ortalama Mutlak Hata (MAE) değeri 4.094 ve R^2 skoru 0.969 ile diğer algoritmalardan daha iyi performans gösterdiğini ortaya koymaktadır. Bu çalışma, güneş enerjisi üretiminin daha güvenilir bir şekilde tahmin edilmesine katkıda bulunarak yenilenebilir enerji kaynaklarının entegrasyonunu kolaylaştırmakta ve enerji yönetim sistemlerinin optimizasyonu için önemli bir potansiyel sunmaktadır.

Anahtar Kelimeler: *PV Santraller, Güneş Enerjisi Üretim Tahmini, Makine Öğrenmesi, Yenilenebilir Enerji*

1. INTRODUCTION

With growing population and production, energy demand is constantly increasing worldwide. As the environmental impacts of fossil fuels have become increasingly evident, the importance of renewable energy sources has increased. According to data from the Turkish Electricity Transmission Company (TEIAS), Turkey's total electricity generation capacity will reach 114,342 MW by October 2024, of which 67,223 MW will be generated from renewable energy sources [1]. Within the power generated from renewable sources, solar power plants (SPPs) have a total installed capacity of 18,839 MW. Turkey's National Energy Plan, published in 2022, aims to increase solar power by 3,500 MW per year until 2035, bringing the total installed capacity to 52.9 GW [2].

The increase in the number of SPPs makes the production forecast of the energy to be generated from these plants and the safe management of the system critical for grid security. The annual variability in seasonal patterns caused by global warming has made day-ahead production forecasts based on short-term weather predictions increasingly valuable. Additionally, consistent and accurate forecasting not only reduces costs and uncertainties but also prevents potential penalties arising from discrepancies between forecasted and actual production, thereby protecting plant investors from potential financial losses [3]. The generation capacity of SPPs is highly dependent on weather conditions; the amount of generation is affected by various meteorological factors such as temperature, solar radiation, cloudiness, humidity and wind speed. In this context, the accuracy of models used for power generation forecasting relies on an effective assessment of these factors [4].

Machine learning models are playing an increasingly important role in forecasting renewable energy production. By predicting fluctuations in power generation, these models contribute to maintaining grid stability, optimizing costs and efficient integration of renewable energy sources. In particular, production forecasting of photovoltaic (PV) systems has become critical as this technology has become more accessible and widespread. Machine learning in solar power generation forecasting offers high accuracy rates and better performance can be achieved by comparing or combining different models. For example, in one study, an approach using the outputs of deep learning models, Artificial Neural Networks (ANN) and Long Short Term Memory (LSTM), as inputs for the XGBoost model improved the R^2 value by 10%-12% compared to other methods [5]. Machine learning models are widely used for short-, medium- and long-term forecasting of PV power generation, and the performance of different algorithms in power generation forecasting is continuously evaluated with data generated under various conditions.

In his study, Tsai (2023) emphasized that models developed with the integration of meteorological data give more successful results in short-term PV generation forecasts, while the accuracy rate of hybrid models is close to 90% [4]. In a study conducted in Morocco, six different machine learning algorithms, namely Support Vector Regressor (SVR), ANN, Decision Trees (DT), Random Forest (RF), Generalized Additive Model (GAM) and XGBoost, were trained using solar energy production data and meteorological parameters and the effectiveness of ANN in modeling complex and non-linear relationships was emphasized [6].

In the study where a deep and machine learning-based ConvLSTM1D model was proposed for the prediction of residential-scale PV generation, models such as XGBoost, RF, SVR, Multilayer Perceptron (MLP) and LSTM were evaluated, and it was stated that RF and ConvLSTM1D outperformed the others [7]. In a similar study, an LSTM-based model was found to perform better than other conventional methods for energy production prediction in PV systems [8].

In a study where a total of 64 different machine learning algorithms were used to predict the energy production of a PV system on a university campus in Manchester to reduce the carbon footprint of buildings, it was reported that the RF algorithm provided higher accuracy compared to other methods and was an effective tool in energy management [9]. AlShafeey et al. (2021), using Multiple Regression (MR) and ANN techniques with three years of PV generation data, examined how the choice of input data used in PV power generation forecasting affects model performance and stated that artificial neural networks provide higher accuracy than MR models regardless of the input method [10].

In another study investigating the impact of different meteorological data on PV generation forecasting, Lasso Regression, Support Vector Machines (SVM), RF and Linear Regression models were used on three years of meteorological and PV generation data in the USA and the Netherlands, and it was stated that temperature, humidity and cloud cover are the most critical factors in energy generation forecasting [11]. Mahmud et al. (2021) compared various machine learning algorithms for PV power generation forecasting in the Alice Springs region of Australia and examined machine learning methods for short and long-term forecasts to reduce fluctuations in generation. At the end of their study, they reported that the RF algorithm provided the highest accuracy in PV generation forecasting and data normalization significantly improved the forecasting performance [12].

Nicoletti and Bevilacqua [13], aiming to facilitate energy production forecasts of individual PV panel users, developed two different feed-forward neural network models using numerical weather forecast data and tested the accuracy of these models with experimental data. In the first model, detailed radiation data were used, while in the second model, more easily accessible general weather data were used. As a result of the study, the models obtained accuracy values of R^2 0.879 and RMSE 10.5%, respectively. Similarly, Buonanno et al. [14] examined the combination of linear models, LSTM, XGBoost and LightGBM for energy forecasting in new PV power generation plants with limited data and found that linear models provide lower error rates than other methods.

Focusing on long-term forecasting of solar energy production, Sedai et al. [15] compared statistical (ARIMA), machine learning (SVR), deep learning (LSTM, GRU, CNN) and ensemble models and concluded that the RF model provided 50% higher accuracy compared to other methods. Similarly, Khadke et al. [16] achieved high accuracy values such as maximum R^2 0.87 and MSE 0.002 with the machine learning-based prediction model they developed using weather data. Andi A. H. Lateko et al. [17], working with data from Zhangbin Industrial Zone in Taiwan, proposed a regression-based ensemble method and achieved a 20%

improvement in MRE compared to a single RF model. These studies show that different types of data and modeling techniques play a critical role in PV power generation forecasting.

Voyant et al. (2017) evaluated various machine learning techniques, including ANN, SVM, decision trees, KNN, and hybrid methods. Their study concluded that single-model approaches such as SVR, ANN, KNN, regression trees, boosting, and random forests produced better predictions compared to classical regression models. They also noted that further research is needed for the KNN algorithm to serve as an alternative [18].

Su et al. (2019) evaluates ten machine learning algorithms, including six neural networks and four intelligent methods, for short-term photovoltaic forecasting. A novel hybrid prediction approach is introduced, integrating the top-performing models to improve forecasting accuracy. Among the tested methods, the Nonlinear Autoregressive Neural Network with Exogenous Inputs (NARXNN) achieved the highest performance among neural networks, while Random Forest (RF) stood out among intelligent algorithms. Their hybrid model demonstrated superior accuracy, achieving the lowest overall normalized RMSE (nRMSE) of 6.74%, underscoring its effectiveness in enhancing forecasting precision [19].

This study aims to develop a forecasting model using past production and weather data through machine learning algorithms and to predict energy production based on future weather forecasts using this model. The main objective of this study is to improve the accuracy of power generation forecasting by improving commonly used machine learning models with hyperparameter optimizations and to identify the most appropriate modeling approach. By presenting an innovative approach to improve the accuracy of machine learning methods in solar photovoltaic (PV) generation forecasting, this study addresses a critical need, especially in countries like Turkey that are rapidly expanding their renewable energy capacity. Generation forecasting of PV systems is of great importance for both maintaining grid stability and optimizing energy management. Although there are many studies in the literature where accuracy is improved by integrating meteorological data, combining different machine learning algorithms and using hybrid models, the capacity of existing models to adapt to specific meteorological conditions and geographical constraints is often limited. In this context, we aim to improve the accuracy of power generation forecasting with machine learning models developed using real-time PV generation data collected in Turkey and detailed meteorological parameters. The study makes a significant contribution to the literature by not only optimizing short- and medium-term forecasts, but also by comparing the performance of different models and providing an applicable framework for energy management.

2. MATERIAL and METHODS

The flowchart of the system designed in this study, which uses machine learning methods for accurate prediction of energy production in photovoltaic solar power plants, is given in Figure 1. After data collection, data preprocessing was performed, data from different sources were combined, tested with machine learning models and the most appropriate prediction model was determined by comparing the results obtained.

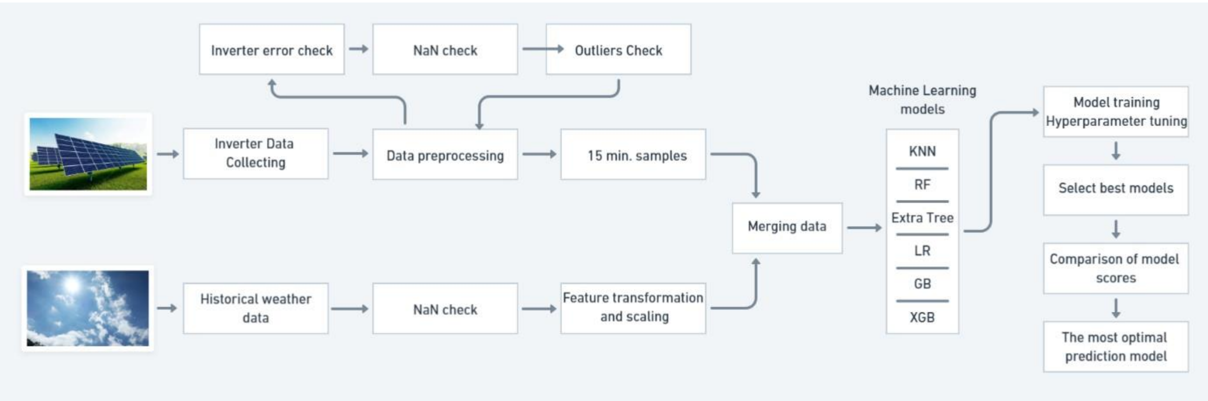

Figure 1. Flowchart of the designed system

2.1. Dataset and Preprocessing

The production data used in this study was obtained from a SMA brand inverter device with a capacity of 1 MW. From the raw inverter data, the records with state codes 512 and 513, which contain the states in which the inverter is operating properly, were selected and the inverter dataset was created by checking for missing data. The maximum value that the inverter can generate in 5 minutes was calculated by equation (1).

$$(1000/60) \times 5 \cong 83{,}33 \ (kW) \quad (1)$$

In addition, weather data with 15-minute periods were obtained from Solcast (weather forecasting service). These data include meteorological parameters such as solar radiation, temperature and wind speed. Table 1 shows the weather parameters used with their descriptions and units.

Erroneous and larger than calculated production values were identified and linear interpolation methods were applied on these values. In order to harmonize the production data with the weather data, the production data were converted into 15-minute periods and the values produced in each period were summed.

Weather parameters include data with different units. For the data with angular features, sine and cosine transformations were applied to enable the model to better learn the periodic information and new features were added to the dataset with these transformations. In addition, all other numerical features were normalized by applying the StandardScaler method.

The successful application of machine learning models is highly dependent on the quality and accurate preprocessing of the data. These processes directly affect the accuracy and performance of energy production forecasting models.

Table 1. Weather dataset parameters

Parameter	Description	Unit
air_temp	Air temperature measured at a height of 2 meters	°C
albedo	Indicates how much sunlight the surface reflects	value between 0-1
azimuth	The azimuth of the sun, measured clockwise from the north	Degree (°)
cloud_opacity	Indicates how much sunlight is blocked by clouds	% (percent)
dewpoint_temp	Indicates the temperature point at which dew starts to form	°C
dhi	Diffuse Horizontal Irradiance, diffuse solar radiation reaching the surface horizontally	W/m^2
dni	Direct Normal Irradiance, solar radiation reaching the surface directly	W/m^2
ghi	Global Horizontal Irradiance, total solar radiation reaching the surface horizontally	W/m^2

gti	Global Tilted Irradiance, total solar radiation reaching a tilted surface	W/m²
precipitation_rate	Rainfall rate, the amount of precipitation that falls in a given period of time	mm/hour
relative_humidity	Refers to the relative humidity at 2 meters height	% (percent)
surface_pressure	Surface pressure, atmospheric pressure value	hPa (hektopascal)
wind_speed_10m	Wind speed, measured at a height of 10 meters	m/s
zenith	The zenith angle of the sun, the angle formed by the sun's rays with the earth's surface	Degree (°)

Proper data preparation facilitates the model learning process, reduces the risk of over-learning and helps the model to make more reliable predictions. Standardized data contributes to more effective optimization of weights during model training. Furthermore, well-prepared datasets for energy forecasting significantly improve the accuracy of forecasting [12].

After the initial preprocessing on the inverter and weather datasets, the data were merged according to time columns, eliminating the time differences between the two datasets and making them ready for model training. Following the data preprocessing step, the main dataset, formed by merging the inverter and weather datasets, comprises 70,127 rows of data spanning from 2022-04-01 04:30:00 to 2024-03-31 13:45:00, with 15-minute intervals. This dataset is split into 80% for training and 20% for testing.

2.2. K-Nearest Neighbors

The K-Nearest Neighbors (KNN) algorithm is a machine learning method used in classification and regression problems that groups data points based on the characteristics of neighboring samples. It is especially preferred in renewable energy forecasting, where historical data is used to predict future trends. KNN is an effective method for solar and wind energy forecasting due to its simple structure and high accuracy rates in various data sets. In one study, KNN was used to predict solar and wind power generation, and 90% accuracy was achieved, especially in daily forecasts. It is stated that these results will provide a great advantage in terms of increasing the security of energy grids and reducing costs [20].

2.3. XGBoost

XGBoost (eXtreme Gradient Boosting) is a powerful machine learning algorithm based on gradient boosting decision trees. With its fast computational capability and robustness against overlearning, it provides effective forecasting results on large data sets. XGBoost is especially preferred for PV power generation forecasts. In a study, the XGBoost model was used in combination with time series such as ANN and LSTM for solar energy forecasting with an R^2 score of 98% and a low error rate compared to other algorithms [5].

2.4. Random Forest

Random Forest is a powerful model that combines multiple decision trees to achieve both high accuracy in energy prediction and reduce the risk of overlearning. Each tree is trained on a random subset of the dataset and the forecast is made by averaging these trees. RF is widely used in solar and wind power generation forecasting due to its high accuracy and robustness to overlearning. Different tree models along with RF have been tested for short-term wind power forecasting and found to be robust techniques [23].

2.5. Extra Trees

Extra Trees is similar to Random Forest, but differs in that it chooses the split points in each decision tree completely at random. This randomization strategy introduces more diversity into the model, reducing the risk of overlearning and resulting in highly accurate forecasts. The Extra Trees algorithm has been observed to provide consistent forecast performance, especially under different meteorological conditions and across a variety of data sets [21]. In addition, it has been reported that this model has a faster computational process compared to other tree-based methods and improves forecast accuracy.

2.6 Gradient Boosting

Gradient Boosting, as an ensemble model of weak learners, improves overall performance by learning and reducing the errors of the previous model at each iteration. Although it is generally highly accurate, the computational cost can be high. However, it can be highly effective on time series data used in solar energy forecasting. The Gradient Boosting model was evaluated together with other decision tree models and it was observed that the ensemble model provided a higher prediction accuracy compared to other boosting algorithms used alone [22].

2.7. Lineer Regresyon

This model allows the prediction of a dependent variable based on a given independent variable or set of variables. It is widely used in power generation forecasting, such as PV power forecasting, due to its simplicity and easy applicability. It offers high performance and improves the interpretability of forecasts, especially when the linear relationship between the dependent and independent variables is evident. In a study conducted by H. Sarper et al. in 2021, linear regression models were used to predict the daily energy production of three different PV systems. Trained with four years of data, it was emphasized that the model can provide additional benefits when combined with other advanced forecasting methods [24].

2.8. Evaluation Parameters

The metrics used to evaluate the performance of machine learning models - Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-Squared (R^2) are critical for measuring the accuracy and bias of predictions. Each of these metrics helps determine which model is better by analyzing different aspects of model performance.

2.8.1. Mean Squared Error

MSE is the mean of the squares of the differences between predicted values ($\hat{y}$) and actual values (y). Its ability to highlight large errors is useful in identifying models with extreme bias. It measures the overall error of the model, but the results are in terms of the square of the prediction unit.

Mathematically, it is expressed as shown in function 2.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (2)$$

2.8.2. Root Mean Squared Error

It is calculated by taking the square root of the MSE. This metric allows the error to be expressed in the original units. It presents the results in the same units as the predicted variable and is often preferred when evaluating prediction accuracy. It is mathematically defined in function 3.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

2.8.3. Mean Absolute Error

It is the average of the absolute values of the differences between predicted and actual values. Although it is not as sensitive as MSE or RMSE in terms of assessing the magnitude of error, it is less sensitive to large errors because it measures errors directly in absolute value. It is therefore highly interpretable and not affected by extreme deviations. It is expressed mathematically in function 4.

$$MAE = \frac{1}{n} \sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

2.8.4. R-Squared

It is a measure of the proportion of a model's independent variables that can explain the dependent variable. It expresses the overall performance of the model with a single ratio and takes a value in the range 0-1. The closer the results are to 1, the higher the explanatory power of the model. Mathematically, it is calculated as in function 5 [6].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5}$$

3. RESULTS and DISCUSSION

In this study, XGBoost, Extra Trees, k-Nearest Neighbors (KNN), Gradient Boosting, Random Forest and Linear Regression machine learning models are evaluated and their performances are compared for the prediction of photovoltaic (PV) energy generation. Hyperparameter optimization plays a crucial role in determining the performance of machine learning models. Effectively managing this process significantly enhances model accuracy while also reducing training costs, providing an advantage [25]. In model training, the most appropriate hyperparameter values were determined for each algorithm with the RandomizedSearchCV method. The best tested hyperparameter values are presented in Table 2. Techniques such as data normalization, sine and cosine transformations increased the learning capacity of the models and improved the overall accuracy.

Table 2. Best hyperparameter values of the models used

Model	Hyperparameter	Values
KNN	weights	distance
	p	1
	n-neighbors	5
	leaf-size	50
	algorithm	kd_tree
XGBoost	subsample	0.6
	reg-lambda	2
	reg-alpha	0.1
	n-estimators	300
	min-child-weight	1
	max-depth	9
	learning-rate	0.05
	gamma	0.1
	colsample-bytree	0.8
Gradient Boosting	subsample	1.0
	n-estimators	500
	min-sample-split	2
	min-samples-leaf	2
	max-depth	9
	loss	squared_error

	learning-rate	0.1
	alpha	0.95
	validation_fraction	0.2
Random Forest	n-estimators	500
	min-samples-split	5
	min-samples-leaf	1
	max-features	sqrt
	max-depth	20
Extra Trees	n-estimators	100
	min-samples-split	2
	min-samples-leaf	1
	max-features	log2
	max-depth	30
Linear Regression	-	-

This optimization process contributed to a significant increase in accuracy and reduced prediction errors, especially in the KNN model. The model scores obtained using the test data are presented in Table 3 to evaluate in detail the performance of the machine learning algorithms in energy production forecasting. In this table, MSE, RMSE, MAE and R^2 metrics are given for each model. It is seen that the KNN model has the lowest error rates and the highest R^2 score compared to the other models. Especially the low RMSE value shows that the production values predicted by KNN are very close to the actual values and that it strongly models the relationship between the dependent variable and the independent variables. KNN algorithm stands out with its high accuracy rates. The scores obtained for the KNN model are given in table3 as follows: MSE value 112.282, RMSE value 10.596, MAE value 4.094 and R^2 score 0.969.

Table 3. Prediction scores of the models using test data

Model	MSE	RMSE	MAE	R2
KNN	112.282	10.596	4.094	0.969
Gradient Boosting	125.604	11.207	4.344	0.965

XGBoost	131.177	11.453	4.398	0.964
Extra Trees	148.614	12.190	4.611	0.959
Random Forest	150.619	12.272	4.580	0.958
Linear Regression	197.996	14.071	6.699	0.94

One-day production prediction scores with test data are presented in Table 4. Similarly, the KNN algorithm is found to be more accurate than the other algorithms, although it produces lower scores with a small deviation from its score in Table 3. These tables show the performance advantages of different models in a comparative perspective.

A comparison of the actual and predicted values using test data is presented in Figure 2 to visualize the success of the model predictions. The graph shows that while acceptable predictions are obtained in the tested models, the actual and predicted values of the KNN model are quite close to each other, the deviations are minimal and the model generally predicts the data points accurately. This suggests that the model not only improves the overall forecast accuracy but also effectively models production fluctuations.

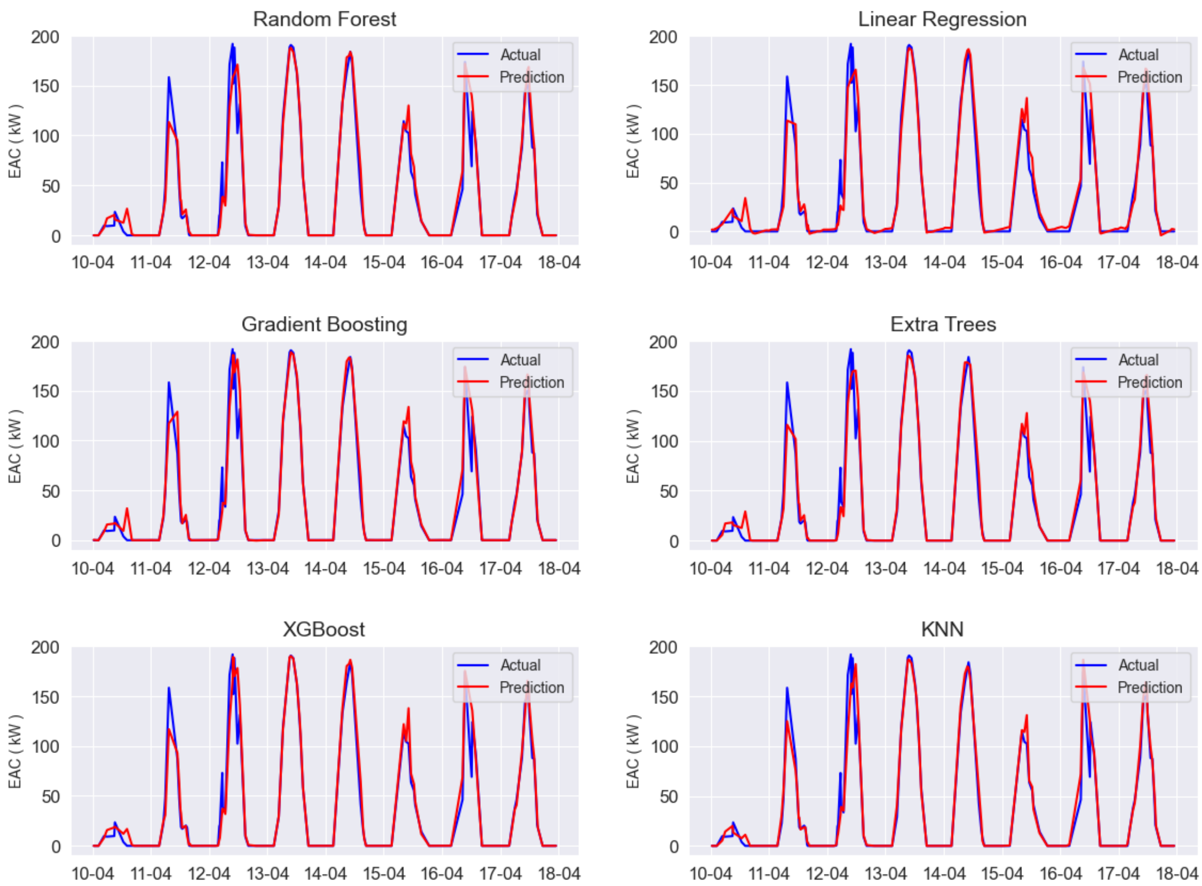

Figure 2. Actual and predicted value comparison graph with test data

This study has showed the potential of machine learning methods in solar power generation forecasting and has shown that the KNN model in particular stands out for its high success rate in this field. The findings provide a valuable framework for optimizing energy management and enhancing grid stability by enabling reliable energy forecasts for PV systems. Future studies may offer the opportunity to further generalize these results with larger datasets and testing in different geographical regions.

This study examines the performance of machine learning models for the prediction of photovoltaic power generation using generation data collected from a 1 MW solar power plant over a two-year period and meteorological data obtained from the Solcast API service. In the study, the effects of important meteorological parameters such as temperature, irradiance, humidity, wind speed on power generation forecasting are considered and various machine learning algorithms are tested. The KNN algorithm showed the best performance compared to other algorithms with 96% R^2 score, 112.282 MSE, 10.596 RMSE and 4.094 MAE. The quality of the data used plays an important role in the performance of machine learning models. In this study, a careful data preprocessing process was applied using two years of datasets with detailed parameters and the accuracy of the models was improved. In particular, harmonization of production and weather data, modeling of periodic features with angular transformations and other scalings have contributed significantly to forecast accuracy. Moreover, the hyperparameter optimization further improved the overall performance of the models.

Similarly, in a study focused on wind farm production forecasting, the KNN algorithm was proven to be a robust predictive model for short-term wind power calculations, achieving an R^2 score of 97% [23]. In another study comparing 24 different machine learning models, the high sensitivity of the KNN algorithm to hyperparameters was emphasized, with findings indicating that proper optimizations allowed it to outperform many other models [21]. In a study on radiation estimation for solar power plants, Uğuz et al. (2019) tested ANN, Multiple Linear Regression (MLR), and KNN algorithms. As a result of their research, they concluded that the ANN algorithm produced the best accurate estimates with an R^2 score of 0.979 [26].

In conclusion, the findings revealed the potential of KNN in solar power generation forecasting models, especially in providing reliable forecasts. Such forecasting models offer important contributions for more efficient management of energy systems and integration of renewable energy sources.

4. CONCLUSION

This study presents an in-depth analysis of machine learning algorithms for solar power plant production forecasting, utilizing an original dataset comprising real-time production data and comprehensive meteorological parameters collected from a solar power plant located in Denizli, Turkey. The uniqueness of the dataset provides a significant contribution by reflecting real field conditions and enabling an objective evaluation of the performance of various algorithms, thereby enhancing the accuracy of prediction systems. The models examined included XGBoost, Extra Trees, KNN, Gradient Boosting, Random Forest and Linear Regression. Among these models, the KNN model performed the best with 96% R^2 score, 112.282 MSE, 10.596 RMSE and 4.094 MAE. These results demonstrate the effectiveness of KNN in providing accurate and reliable predictions. The research contributes to the importance of accurate PV generation forecasting in the field of renewable energy, especially in rapidly growing markets such as Turkey. Improving forecast accuracy through hyperparameter optimizations and model comparisons provides a robust framework for optimizing energy management systems and ensuring grid stability. The integration of real-time meteorological data and comparative analysis of machine learning models provide valuable scientific contributions towards the development of renewable energy integration. This study not only improves short and medium-term forecasts, but also paves the way for future research on hybrid and adaptive models that adapt to different meteorological and geographical conditions.

Çıkar Çatışması Beyanı

Yazarlar arasında çıkar çatışması yoktur.

Araştırma ve Yayın Etiği Beyanı

Çalışma, araştırma ve yayın etiğine uygundur.

REFERENCES

[1] Enerji Ajansı, "Türkiye’nin Kurulu Gücü (Ekim 2024)," 2024. [Online]. Available: https://enerjiajansi.com.tr/turkiyenin-kurulu-gucu/. [Accessed: Nov. 21, 2024].

[2] T.C. Ministry of Energy and Natural Resources, *Turkey National Energy Plan*, Ankara, Turkey, 2022.

[3] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," Solar Energy, vol. 136, pp. 78-111, 2016.

[4] W.-C. Tsai, C.-S. Tu, C.-M. Hong, and W.-M. Lin, "A review of state-of-the-art and short-term forecasting models for solar PV power generation," *Energies*, vol. 16, no. 14, p. 5436, 2023.

[5] W. Khan, S. Walker, and W. Zeiler, "Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach," *Energy*, vol. 240, p. 122812, 2022.

[6] Y. Ledmaoui, A. E. Maghraoui, M. E. Aroussi, R. Saadane, A. Chebak, and A. Chehri, "Forecasting solar energy production: A comparative study of machine learning algorithms," *Energy Reports*, vol. 10, pp. 1004–1012, 2023.

[7] S. Cantillo-Luna, R. Moreno-Chuquen, D. Celeita, and G. Anders, "Deep and machine learning models to forecast photovoltaic power generation," *Energies*, vol. 16, no. 10, p. 4097, 2023.

[8] D. K. Dhaked, S. Dadhich, and D. Birla, "Power output forecasting of solar photovoltaic plant using LSTM," *Green Energy and Intelligent Transportation*, vol. 2, no. 5, p. 100113, 2023.

[9] C. Scott, M. Ahsan, and A. Albarbar, "Machine learning for forecasting a photovoltaic (PV) generation system," *Energy*, vol. 278, p. 127807, 2023.

[10] M. AlShafeey and C. Csáki, "Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods," *Energy Reports*, vol. 7, pp. 7601–7614, 2021.

[11] T. AlSkaif, S. Dev, L. Visser, M. Hossari, and W. van Sark, "A systematic analysis of meteorological variables for PV output power estimation," *Renewable Energy*, vol. 153, pp. 12–22, 2020.

[12] K. Mahmud, S. Azam, A. Karim, S. Zobaed, B. Shanmugam, and D. Mathur, "Machine learning-based PV power generation forecasting in Alice Springs," *IEEE Access*, vol. 9, pp. 46117–46128, 2021.

[13] F. Nicoletti and P. Bevilacqua, "Hourly photovoltaic production prediction using numerical weather data," *Energies*, vol. 17, no. 2, p. 466, 2024.

[14] A. Buonanno et al., "Machine learning and weather model combination for PV energy forecasting," *Energies*, vol. 17, no. 9, p. 2203, 2024.

[15] A. Sedai et al., "Performance analysis of statistical, machine learning, and deep learning models in long-term forecasting of solar power production," *Forecasting*, vol. 5, no. 1, pp. 256–284, 2023.

[16] S. Khadke et al., "Predicting active solar power with machine learning and weather data," *Materials Circular Economy*, vol. 5, no. 15, 2023.

[17] A. A. H. Lateko, H.-T. Yang, and C.-M. Huang, "Short-term PV power forecasting using a regression-based ensemble method," Energies, vol. 15, p. 4171, 2022.

[18] C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[19] D. Su, E. Batzelis, and B. Pal, "Machine learning algorithms in forecasting of photovoltaic power generation," in *2019 International Conference on Smart Energy Systems and Technologies (SEST)*, 2019, pp. 1-6.

[20] T. Ahmad, S. Manzoor, and D. Zhang, "Forecasting high penetration of solar and wind power in the smart grid environment using a robust ensemble learning approach for large-dimensional data," *Sustainable Cities and Society*, vol. 75, p. 103269, 2021.

[21] D. Markovics and M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," *Renewable and Sustainable Energy Reviews*, vol. 161, p. 112364, 2022.

[22] A. Abdellatif et al., "Forecasting photovoltaic power generation with a stacking ensemble model," Sustainability, vol. 14, no. 17, p. 11083, 2022.

[23] U. Singh, M. Rizwan, M. Alaraj, and I. Alsaidan, "A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments," *Energies*, vol. 14, p. 5196, 2021.

[24] H. M. I. and M. L. Sarper, "Prediction of daily photovoltaic energy production using weather data and regression," *Journal of Solar Energy Engineering-Transactions of the ASME*, 2021.

[25] J. Wu, X. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S. Deng, "Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization," *Journal of Electronic Science and Technology*, vol. 17, pp. 26-40, 2019.

[26] S. Uğuz, O. Oral, and N. Çağlayan, "PV Güç Santrallerinden Elde Edilecek Enerjinin Makine Öğrenmesi Metotları Kullanılarak Tahmin Edilmesi," *International Journal of Engineering Research and Development*, vol. 11, no. 3, pp. 769–779, 2019.