BMBAD BJESR | Mühendislik Bilimleri ve Araştırmaları Dergisi
Journal of Engineering Sciences and Researches

Estimating Solar Energy within the scope of environmental factors by the Neural Network Algorithm

Sinir ağı algoritması ile çevresel faktörler kapsamında Güneş Enerjisinin tahmin edilmesi

[1]Yasemin AYAZ ATALAN

[1]Yozgat Bozok University, Faculty of Engineering and Architecture, 66000, Yozgat, Türkiye

[1]yasemin.ayaz@yobu.edu.tr

Araştırma Makalesi/Research Article

ARTICLE INFO

Article history

Received : 17 October 2023
Accepted : 12 December 2023

Keywords:

Machine Learning, Neural Network Algorithm, Solar Energy, Environmental Factors

ABSTRACT

The efficiency of solar energy systems requires a complicated forecasting process due to the variability of sunlight and environmental conditions. Among environmental factors, cloud coverage (% range), temperature (^{0}C), wind speed (Mph), and humidity (%) variables were taken into account in this study. Neural networks (NN), which are machine learning (ML) algorithms with a flexible structure that can define complex relationships and process large amounts of data for solar energy prediction, were used in this study. The NN algorithm showed a high performance, with mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and R-squared (R^2) values calculated as 0.019, 0.139, 0.053, and 0.977, respectively. This study emphasized that solar energy predictions made with the NN algorithm, considering environmental factors, are an essential tool that helps use solar energy systems more efficiently and sustainably.

MAKALE BİLGİSİ

Makale Tarihleri

Gönderim : 17 Ekim 2023
Kabul : 12 Aralık 2023

Anahtar Kelimeler:

Makine Öğrenimi, Sinir Ağı Algoritması, Güneş Enerjisi, Çevresel Faktörler

ÖZET

Güneş enerjisi sistemlerinin verimliliği, güneş ışığının ve çevre koşullarının değişkenliği nedeniyle karmaşık bir tahmin süreci gerektirmektedir. Bu çalışmada çevresel faktörlerden bulut kapsamı (% aralık), hava sıcaklığı (0C), rüzgar hızı (Mph) ve bağıl nem (%) değişkenleri dikkate alınmıştır. Bu çalışmada güneş enerjisi tahmini için karmaşık ilişkileri tanımlayabilen ve büyük miktarda veriyi işleyebilen esnek yapıya sahip makine öğrenmesi (ML) algoritmalarından sinir ağı (NN) kullanılmıştır. NN algoritması, ortalama kare hatası (MSE), kök ortalama kare hatası (RMSE), ortalama mutlak hatası (MAE), ve R-kare (R^2) değerleri sırasıyla 0,019, 0,139 0,053 ve 0,977 olarak hesaplanarak yüksek bir performans göstermiştir. Bu çalışmada, çevresel faktörler dikkate alınarak NN algoritması ile yapılan güneş enerjisi tahminlerinin, güneş enerjisi sistemlerinin daha verimli ve sürdürülebilir kullanılmasına yardımcı olan önemli bir araç olduğu vurgulanmıştır.

ORCID ID: [1]0000-0001-7767-0342

1. INTRODUCTION

Solar energy, a renewable energy source, is growing in significance within the energy production sector and offers substantial potential for environmental sustainability [1]. Solar energy is an environmentally friendly energy source [2]. Greenhouse gases produced by burning fossil fuels cause serious problems such as global climate change and air pollution [3] [4]. On the other hand, solar energy can produce clean electricity that does not harm the environment [5]. Therefore, solar energy supports environmental protection efforts by helping reduce the carbon footprint [6].

The source of solar energy is unlimited and abundant worldwide [7]. The sun sends a tremendous amount of energy to the Earth every day, and only a tiny part of this energy is used by humans [8]. This is a great advantage in terms of the sustainability of energy supply. Solar energy provides long-term energy security compared to energy production based on limited resources such as fossil fuels [9]. Solar energy is also an economically competitive option. With technological advances and large-scale production, solar panels and energy storage systems are becoming more accessible and cost-effective [10]. This helps individuals, businesses, and governments reduce energy costs and achieve energy independence by using solar energy. For all these reasons, solar energy is becoming an increasingly preferred energy source worldwide. Being clean, unlimited, and economically attractive, solar energy increases its role in energy production and becomes an essential driving force in achieving environmental sustainability goals [11].

Solar energy and artificial intelligence (AI) have a strong relationship between the two fields, creating significant collaboration in the energy sector [12]. AI is an essential tool to increase the efficiency of solar energy structures, make predictions, optimize maintenance, and manage energy production more sustainably [13]. In particular, AI-based forecasting models can predict solar energy production more precisely by analyzing sunlight levels, weather conditions, and other factors that affect energy production [14]. It can also increase the operational continuity of systems by using sensors and AI to monitor the maintenance of solar panels and detect failures in advance. In this way, solar energy and AI support the energy transition by providing a more efficient and reliable path for clean and sustainable energy production [15].

The relationship between solar energy and machine learning (ML) algorithms creates a vital collaboration to increase the efficiency of solar energy plants, optimize energy production, and make energy management more sustainable [1]. ML algorithms analyze various variables that affect solar energy production, offering the ability to predict sunlight levels, weather conditions, location of panels, and other environmental factors [16]. These forecasts provide valuable information to energy producers and distributors to optimize the performance of solar energy systems and better manage energy demand. Additionally, ML can be used to monitor the maintenance of solar systems, detect failures in advance, and improve system efficiency [17]. In this way, solar energy and ML contribute to more effective, clean, and sustainable energy production, increasing energy efficiency and promoting environmental sustainability.

Artificial neural networks (ANNs), a subset of ML algorithms, have a significant impact on the field of solar energy. The complexity of solar energy systems and environmental variability make it difficult to predict solar energy production [18]. The NN algorithm offers an effective tool for predicting and optimizing solar energy production with its ability to process large amounts of data and model complex relationships [19]. This algorithm, which can include environmental factors such as sunlight levels, weather conditions, and cloud cover, as well as system features such as the location and tilt of solar panels, allows for more precise predictions. This offers a solar energy system that is both more dependable and efficient for energy producers, grids, and consumers. The NN algorithm emerges as a crucial instrument for enhancing the efficiency of solar energy systems and advancing the production of clean energy [20].

The NN algorithm should be used for solar energy because solar energy systems operate complexly under the influence of environmental variations, and traditional methods cannot fully address the efficiency and predictability of these systems [21]. With its ability to process large amounts of data, the NN algorithm can more precisely predict solar energy production by considering sunlight levels, weather conditions, cloud cover, and other environmental factors [19]. The NN algorithm also helps optimize energy production by modeling system features such as the location and tilt of solar panels. This situation contributes to reducing energy costs while increasing the efficiency of solar energy systems, promoting clean and sustainable energy production. Therefore, the NN algorithm may be the key to a more reliable and efficient future in the solar energy industry. A study reveals that a NN-based tool developed to estimate the amount of solar energy produced by photovoltaic generators addresses the control complexity in solar energy systems and has the potential to optimize future energy production [18]. Another study created complex non-linear models using neural networks, emphasizing the importance of using a range of predictions that include uncertainty rather than typically a single prediction for renewable energy predictions [22].

One of the most important contributions of this study to the literature and the industry is to create a model by creating an NN model in ML algorithms, which is reliable among prediction models. This study, which considers weather parameters, has resulted in the emergence of a manageable system in the installation phase or in calculating the energy amounts of installed solar power plants. The ability to learn complex relationships and patterns by analyzing large amounts of meteorological data in this study is provided by the NN algorithm. Thanks to their ability to analyze real-time data, NN algorithms can quickly adapt to changes in weather conditions and

allow solar energy systems to be managed more effectively. This study is planned to contribute to the development of more precise and reliable prediction models in the field of solar energy, leading to significant advances in energy planning and sustainable energy use.
This study consists of four main sections. Information about the importance of solar energy, one of the renewable energy types, its advantages, the factors that affect it, and the place of the method used in this study in the literature are discussed in the first part of the research. The second section provides details on the methods employed in the study, followed by a discussion of the study's statistical and numerical findings in the third section. The final section presents overarching remarks regarding the scientific contributions of this research.

2. METHODOLOGY

This study aims to create an NN model from ML algorithms, taking into account environmental factors, to calculate the amount of solar energy production within the scope of renewable energy. ML makes significant contributions to making predictions about future solar radiation and weather conditions by analyzing various meteorological data. The most important reason for using this method is that ML algorithms allow more effective planning and management of solar energy systems. Solar energy is an energy source that can fluctuate depending on atmospheric conditions, so predicting how long and how intense the sun's rays will be is critical for optimizing energy production and increasing the efficiency of solar energy systems. By processing large data sets, ML reveals complex patterns, which helps predict solar energy production more accurately. Publicly available data on solar energy amounts and environmental inputs were used in the present research [23]. The flow chart of the techniques used in the methodology of this study is visualized in Figure 1.

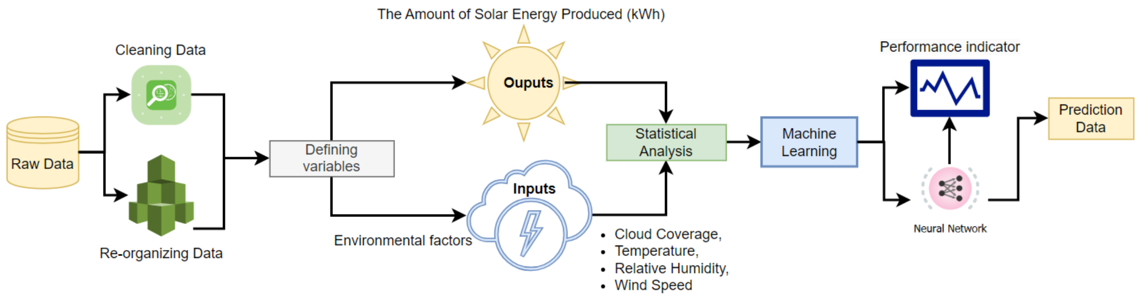

Figure 1. The visual representation outlining the methodological process.

This study has a three-step method. The first step involves pre-examination of the study data. The second step of the method includes defining variables and statistical analysis. In the last step of the method, prediction data was obtained by running the model of the ML algorithm.

2.1. Data of Study and Descriptive Statistics

This research treats the quantity of solar energy as the dependent variable, while it characterizes the independent variables as environmental factors influencing the solar energy amount. The statistical properties of **six** variables (Cloud Coverage, Temperature, Relative Humidity, Wind Speed, and Solar Energy) and the results of the data analysis about these variables are displayed in Table 1. The inputs of the study are defined as Cloud Coverage, Temperature, Relative Humidity, Wind Speed, and hour (not included in Table 1) and the output is defined as the amount of solar energy.

Table 1. Descriptive statistical values of data sets regarding the amount of Solar Energy and environmental factors.

Variable	Cloud Coverage (% range)	Temperature (^{0}C)	Relative humidity (%)	Wind speed (Mph)	Solar Energy (kWh)
Sample Size	7536.00	7536.00	7536.00	7536.0	7536.0
Mean	0.41369	16,4300	65,4560	10,111	17,962
Standard Error of Mean	0.00482	0.1220	0.23700	0.0654	0.0105
Standard Deviation	0.41864	10,5950	20,6060	5,6800	0.9151
Variance	0.17526	112,248	424,596	32,267	0.8375
Minimum	0.00000	-18,9300	14,8800	0.0000	0.0000
Maximum	100,000	34.0100	100,000	41,560	46,738
Skewness	0.34000	-0.62000	-0.32000	0.6900	0.0200
Kurtosis	-1.59000	-0.41000	-0.78000	0.6600	-0.9100

The number of observations for each variable is expressed as 7536 in total in the data set. The measures of statistical describe each variable's central tendency, dispersion, skewness, and kurtosis. The Cloud Coverage variable indicates the proportion of cloud cover. The average cloud cover percentage was calculated as 0.41369. The standard deviation value of 0.41864 indicates how spread this distribution is around the data. The skewness value is 0.34, indicating a slight rightward skew in this distribution. Also, the kurtosis value is -1.59, indicating that the distribution has some weighted tails. The temperature variable shows the average temperature value as 16,430. The standard deviation is 10.595, meaning that this distribution has a wide range of data. The skewness is -0.62, indicating that this distribution is slightly skewed to the left, while the kurtosis value is -0.41, showing a flatter distribution.

The relative humidity variable shows the average relative humidity as 65,456. The standard deviation is 20.606; this distribution reflects the degree of spread of the data. The skewness value is -0.32, indicating that the distribution is slightly skewed to the left, while the kurtosis value is -0.78, showing the flatness of the distribution. The average value of the wind speed variable is 10.111. The standard deviation is 5.680, indicating that the wind speed data of this distribution are somewhat variable. The skewness is 0.69, indicating that this distribution is skewed to the right, while the kurtosis value is 0.66, indicating that the distribution is more peaked. The solar energy variable shows an average solar energy value of 17,962. The standard deviation is 0.9151, indicating this distribution has a relatively low variance. The skewness value is 0.02, indicating that this distribution is slightly skewed to the right, while the kurtosis value is -0.91, reflecting the flatterness of this distribution.

In this study, independent and dependent variable data were used in the NN model, one of the ML algorithms, to estimate the amount of solar energy. The rest of this section discusses the criteria that measure the performance of ML, NN algorithm, and ML models.

2.2. Machine Learning (ML)

Machine learning (ML) is a subfield of artificial intelligence (AI) that empowers computer systems to acquire the capability to learn from data and enhance future decisions or predictions based on accumulated experience [24]. ML helps computers perform human-like intelligence functions, with the ability to analyze large amounts of data and discover patterns and relationships, solve complex problems, and make autonomous decisions [25]. ML is assessed within three fundamental categories: supervised, unsupervised, and reinforcement learning [26]. Supervised Learning learns a model on training data and then applies this learned information to new data [27]. This type of learning works with labeled data, meaning there is correct output or label information for each data sample. Classification and regression problems are examples of supervised learning. Unsupervised learning works on unlabeled data and aims to discover hidden structures, patterns, or groupings within the data [28]. It is used to find similarities or differences between samples. Clustering and dimensionality reduction can be given as examples of this category. Reinforcement Learning presents a paradigm in which an agent interacts with an environment and learns by receiving rewards or punishments for performing specific actions against the environment [29]. It is used in game theory, robot control, and automated vehicles.

2.3. Neural Network Algorithm

In the present research, the neural network (NN) algorithm was used, which is within the scope of the supervised learning method. The NN is an essential ML algorithm and an AI model inspired by biological nervous systems [30]. The NN attracts attention with its structure consisting of many connections and layers [31]. These networks can solve complex problems by processing large amounts of data. The NN uses a network structure mainly composed of artificial cells called neurons [32]. These neurons take input data, process them with a weight and activation function, and then produce an output. This process occurs when the input data passes through the layers within the network to deliver results. The network layers connect where information processing and learning occur by transmitting each neuron's output to other neurons [33].

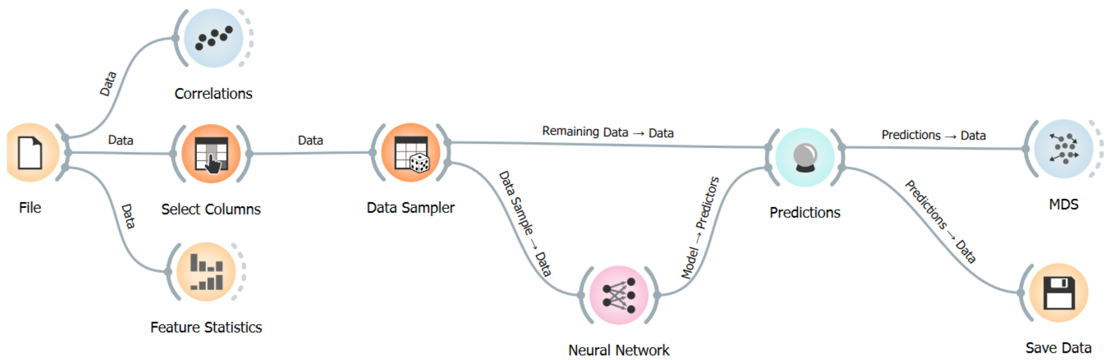

Figure 2. The flowchart of the NN algorithm.

The NN assimilates information through learning processes [34]. There are two primary learning approaches: supervised learning and reinforcement learning [35]. In supervised learning, the network is trained using the data it must learn from, to minimize the disparity between the network's output and the actual values [36]. Reinforcement learning, on the other hand, is a paradigm in which an agent learns by interacting with an environment and receiving rewards or punishments [37]. In this way, the agent learns from its experiences to determine the best course of action. The NNs are used in various application areas [38]. Specifically, it has seen remarkable accomplishments in image recognition, natural language processing, game AI, financial forecasting, medical diagnosis, and numerous other applications [39]. Additionally, a type of NN called deep learning has mainly influenced large and complex data sets, leading to significant advances in the field [40]. NNs have become increasingly used and considered an essential tool in data mining and artificial intelligence. The flowchart of the NN algorithm, one of the ML models developed for this paper, is presented in Figure 2. In this study, the Orange Data Mining program was used to run the NN model, one of the ML algorithms. This program is publicly available and uses Python software. The parameters of the NN algorithm are expressed below:

Model parameters
Model Name: Neural Network
Hidden layers: 200
Activation: ReLu
Solver: Adam
Alpha: 0.003
Max iterations: 100
Replicable training: Yes

2.4. Machine Learning Performance Measurement Criteria

ML performance measurement criteria are used to evaluate a model or algorithm's predictive ability, accuracy, and effectiveness [24]. These criteria help us understand how well or poorly a model works and support decisions made. Error metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or Mean Absolute Error (MAE) determine how close a model's predictions are to actual values [41]. The R-squared (R2) metric also assesses the goodness of fit of a model to the data and the influence of independent variables on the dependent variable [42]. These performance measurement criteria are essential to model development and tuning and comparing different models. They are also important when businesses make decisions and evaluate risks [43]. As a result, ML performance measurement criteria form a fundamental part of our process for assessing model reliability and effectiveness, making our data-driven decisions more robust. Below, the respective performance measurement criteria for ML algorithms are explained.

2.4.1. MSE (Mean Squared Error)

MSE calculates the mean of the squares of variances between the model's predictions and the real values. Smaller MSE values suggest that the model's predictions are closer to the actual values, indicating superior predictive performance [44]. However, using squares can cause more significant errors to be highlighted.

$$MSE = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \tag{1}$$

where, n represents the total number of observations, y_i signifies the estimated values, and y_i denotes the actual values.

2.4.2. RMSE (Root Mean Squared Error)

RMSE is an error metric used to quantify the extent of disparity between a model's predictions and the actual values. It is derived by taking the square root of MSE, making the error more interpretable [45]. Elevated RMSE values signify that the model's predictions exhibit greater deviations from the actual values, indicating lower prediction accuracy. The equation for RMSE is as follows:

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

2.4.3. MAE (Mean Absolute Error)

MAE computes the mean of the absolute variances between the model's predictions and the actual values [44]. This metric directly quantifies the magnitude of forecasting errors and mitigates the influence of substantial errors. Smaller MAE values indicate that the model's predictions, on average, are closer to the actual values. The mathematical equation of the MAE is given below:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \tilde{y}_i|^2 \tag{3}$$

2.4.4. R^2 (R-squared)

R^2 is a statistical measure that assesses how effectively the model aligns with the data. Its values range from 0 to 1, with values closer to 1 indicating a stronger alignment between the model and the data. R^2 signifies the percentage of variance explained, revealing the extent to which independent variables account for the variance in the dependent variable [46]. A higher R^2 value corresponds to a more robust fit of the model to the data and improved prediction accuracy [47]. These metrics are employed to assess and compare the performance of various model types. The suitable metric may vary depending on the data and the context of the application, as each metric evaluates different aspects. The mathematical expression of R^2 is given below:

$$R^2 = \sum_{i=1}^{n} \left[\frac{y_i - \tilde{y}_i}{y_i - \bar{y}_i}\right]^2 \tag{4}$$

Using four different metrics such as MAE, MSE, RMSE, and R^2 to evaluate the performance of ML algorithms provides a more comprehensive evaluation by analyzing various aspects of the model. MAE gives the average of the absolute differences between actual and predicted values and determines the overall accuracy of the model. MSE and RMSE evaluate the model's sensitivity to large errors by taking the mean and square root of the squares of the errors. R^2 measures the ratio of the variance of the independent variables to the dependent variable. R^2 indicates how much of the variance in the data set the model can explain, which is important for assessing the overall fit of the model. When these four measurements are used together, it is possible to evaluate the model's average error, sensitivity to large errors, variance explanation capacity, and overall fit. These multiple measurements help gain a more comprehensive understanding of the model's performance and make improvements. Each measurement provides information from different perspectives and, when used together, paints a more detailed picture of the reliability and applicability of the model.

3. RESULTS AND DISCUSSION

3.1. The Variance Analysis of Variables (ANOVA)

This study considers variables such as Hour, Cloud Coverage, Temperature, Relative Humidity, and Wind Speed as factors affecting solar energy production. Analysis of the variance of these variables showed how solar energy production is associated with these factors. Analysis of variance determines the extent to which these factors explain the variance in solar energy production and how statistically significant each factor is. It reveals the relationships between these factors and the contribution of each factor to solar energy production. For example, factors such as temperature and wind speed can positively or negatively affect solar energy production, and analysis of variance is a crucial tool to measure these effects. Variance analysis of these variables is an essential analytical tool to evaluate their impact on solar energy production and to indicate which factors should be considered to predict energy production. The variance analysis results for the variables considered in this study are shown in Table 2.

Table 2. The results of the ANOVA of variables.

Source	Degree of freedom	Adjusted sums of squares	Adjusted mean squares	F-Value	P-Value
Hour	1	23.34	23.34	2758.37	0.01
Cloud coverage	1	112.66	112.66	13312.14	0.01
Temperature	1	1066.04	1066.04	125964.8	0.02
Relative humidity	1	1639.62	1639.62	193741.1	0.01
Wind speed	1	69.97	69.97	8268.12	0.01

Analysis of variance results for the time variable shows that this variable plays an essential role in the study. The degree of Freedom (DF) value is 1, indicating that this variable participates in the analysis with one degree of freedom. The Corrected Sum of Squares was calculated as 23.34, meaning that the time variable explains the variability in the data. The f-value was 2758.37, and the P-value was 0.01, which shows that the time variable was significantly effective in the analysis. The analysis of variance results for the cloud cover variable is quite impressive. The DF value of this variable is 1, and the Adjusted Sum of Squares value is 112.66. The f-value was calculated as 13312.14, which is relatively high. These findings suggest that cloud cover is a significant variable that predominantly accounts for the variations in the data. Similarly, the analysis results for the temperature variable indicate that it plays a major role in explaining the data's variability. It has a DF value of 1, an Adjusted Sum of Squares value of 1066.04, and a high F-value of 125964.8. These results indicate that the temperature variable had a statistically significant effect on the analysis. The analysis results for the relative humidity variable are pretty impressive. The DF value is 1, and the Adjusted Sum of Squares value is 1639.62. The f-value was calculated as 193741.1, which is relatively high. These results demonstrate that the relative humidity variable is a significant factor that predominantly accounts for the variability in the data and is an important variable. Similarly, the analysis results for the wind speed variable reveal that it plays a significant role in explaining the data's variability. It has a DF value of 1, an Adjusted Sum of Squares value of 69.97, and a high F-value of 8268.12. These results indicate that the wind speed variable has a statistically significant effect on the analysis.

3.2. The Results of the NN Model's Performance Measurement

The NN model, a machine learning algorithm, was employed to estimate the impact of environmental factors on solar energy production. To validate the accuracy of the predictions made by the NN algorithm, performance evaluation metrics were computed. The performance metrics of the NN algorithm are presented in Table 3.

Table 3. The values of performance measurement of the NN model.

Model	MSE	RMSE	MAE	R^2
NN Algorithm	0.019	0.139	0.053	0.977

The MSE value was calculated as 0.019. MSE is an error metric that quantifies the extent of divergence between the model's predictions and the actual values. A lower MSE indicates superior prediction performance, and in this case, the value is quite low. The RMSE value stands at 0.139. RMSE is derived by taking the square root of MSE and provides a more intuitive understanding of the error magnitude. This value is also relatively small, suggesting that the model's predictions are, on the whole, near the actual values. The MAE value is 0.053. MAE calculates the average of the absolute differences between predictions and actual values. This low MAE value implies that the model's predictions generally deviate by approximately 0.053 units. The R2 value was registered as 0.977. R^2 evaluates how well the model fits the data, with values closer to 1 indicating a better fit. In this instance, the R^2 value is notably high, signifying that the model fits the data well, and its predictions are generally very close to the actual values.

In the present paper, the low error and high fit values of all these metrics show that the NN model has an excellent prediction performance. This indicates that the model is a powerful choice in data analysis or prediction tasks and can produce reliable results.

3.3. The Results of Prediction Values

A comparison of the prediction results of the NN algorithm used in the present paper with actual data is discussed in this section. Descriptive statistical data for forecast and actual data are given in Table 4. Comparing the statistical properties of actual data and the values predicted by the NN algorithm, the data show statistically similar properties. The actual value and the value indicated by NN are the same (1506.00). This shows that the total number of data points was estimated correctly. The actual average value is 18476.00, while the value predicted by NN is 18474.00. The estimate is very close, indicating that it accurately captures the center of the data distribution. The standard error values between the actual and predicted values are very similar (0.02). This shows that the mean estimate of the sample data was calculated accurately. There are slight differences between the standard deviation and variance values of the actual data and the predicted values. This indicates that it accurately captures how dispersed the data points are.

Table 4. Descriptive statistical data for forecast and actual data

Variable	Actual Value	Neural Network Value
Total Count	1506.000	1506.000
Mean	18476.00	18474.00
Std. Error of Mean	0.020000	0.020000
StDev	0.930000	0.920000
Variance	0.860000	0.840000
Coefficient of Variance	50.09000	49.69000
Min	0.010000	-0.080000
Q_1	10975.00	10852.00
Q_2	19048.00	19219.00
Q_3	25898.00	25907.00
Max	46738.00	38990.00
Inter Quartile Range	14923.00	15055.00
Skew	-0.040000	-0.050000
Kurt	-0.900000	-0.980000

The estimate is reasonably accurate while there are variances between the observed and projected coefficient of variation. The minimum value of the actual data is 0.01, but the expected value is negative (-0.08). This indicates that there is a slight deviation from the minimum value of the estimate. The maximum data set for the actual value is 46738.00, while the predicted value is lower (38990.00). The quartile values of the real data set and the predicted values are similar. However, there is a slight difference in the interquartile range (IQR). There are similarities between actual and expected asymmetry and kurtosis values. Negative asymmetry and kurtosis indicate the data is close to a normal distribution. In general, the estimates closely approximate the real values, although notable disparities exist in certain minimum and maximum values.

The raw data run on the NN algorithm was divided into 80% for the training and 20% for the testing phases. The number of data used for the testing phase is 1507. The actual data used for the testing phase and the prediction data obtained with the NN algorithm are compared in Figure 3.

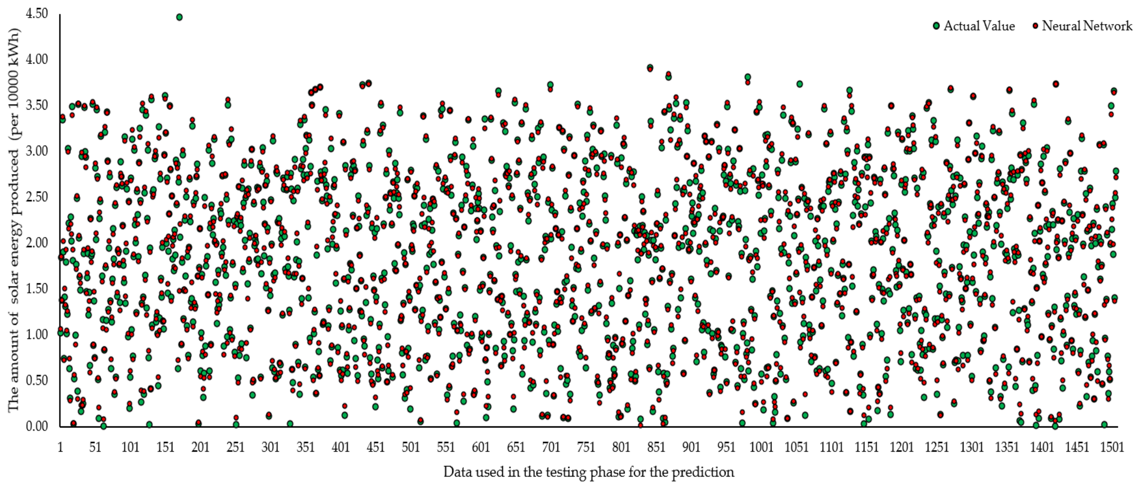

Figure 3. The actual data used for the testing phase and the prediction data.

In one study, researchers highlighted that the SVM model was 27% more accurate than other forecast-based models, considering seven weather forecast measurements [48]. One study used an artificial neural network to predict the amount of solar energy produced and the prediction results showed that the difference between the actual production and the prediction was in the range of 0.5-9%. In this study, the difference between forecast data and actual data was calculated as 0.02% [18]. One of the differences between this study and other studies is the difference in the models used in ML algorithms. Another study emphasized that the deep learning model performed best in ML algorithms to measure the amount of solar energy produced [49]. Another study compared two common methods, artificial neural networks (ANN) and support vector regression (SVR), to predict energy production from a solar photovoltaic (PV) system in Florida 15 minutes, 1 hour, and 24 hours in advance[50]. As a result, both this study and other studies make important contributions to the emergence of other studies.

Multidimensional Scaling (MDS) is a data analysis method used in machine learning and data mining. MDS aims to transform data into a lower dimensional space by preserving the observation similarities of points represented in multidimensional spaces. In this way, it helps to make significant and complex data sets more understandable

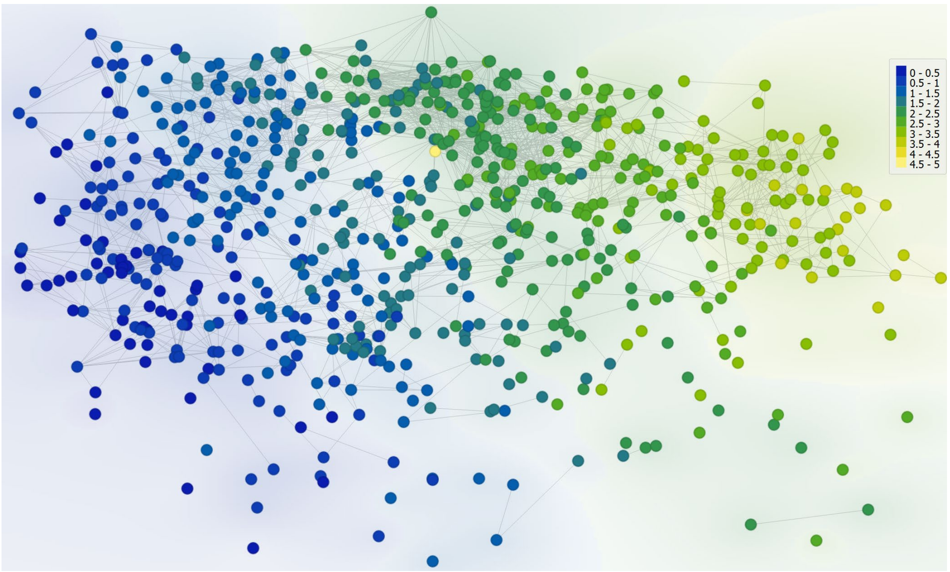

Figure 4. The MDS of the NN algorithm.

and visually representable. MDS facilitates understanding the structure of data sets, recognizing patterns, and discovering relationships between data, especially by using it in data mining, visualization, cluster analysis, and similar applications. This method is considered an important tool, significantly to increase the understandability of multidimensional data and to make data analysis processes more effective. The distribution of the data in the

MSD diagram created for the prediction data of the NN algorithm preferred for this study is shown in Figure 4. The MSD diagram of this algorithm visualizes that the prediction data of the NN algorithm is partially clustered, but the clustering classes are few.
In this study, an NN model from ML algorithms was developed to calculate production amounts for solar energy facilities depending on environmental factors. This study presented a model to estimate the production amounts of solar energy facilities. This study has some limitations. First, information needs to be provided about the physical conditions and equipment of the facility where solar energy amounts are obtained. Secondly, other weather and geographical parameters were not used. Finally, the times of the dates on which the energy amounts were obtained were ignored. As a result, this study presented a model to get predicted values for solar energy, one of the renewable energy types, by considering some restrictions.

4. CONCLUSION

Solar energy is significant among renewable energy sources and is essential for various reasons. Solar energy contributes to electricity production using sunlight and heat and does not produce any by-products or harmful emissions during this process. Additionally, solar energy potential is relatively high worldwide and can increase energy security by reducing dependence on conventional energy sources. Solar power, as a clean, limitless, and eco-friendly energy source, delivers economic advantages while simultaneously helping to achieve sustainability goals in the battle against climate change and energy generation. For these reasons, solar energy is considered an essential component of renewable energy portfolios and is becoming an increasingly preferred option in the energy sector.
This study addressed various environmental factors affecting solar energy production and showed that these factors are associated with solar energy production. By performing variance analysis, the relationships between these factors and the contribution of each factor to solar energy production were revealed. The NN algorithm was utilized to estimate the influence of environmental factors on solar energy production, and model performance metrics were computed to validate these predictions. These metrics indicate the extent of variance between the model's predictions and the actual values. The low MSE and RMSE values achieved in this research demonstrate the model's strong predictive performance, while the high R^2 value signifies a good fit of the model to the data. This implies that the model's predictions are generally in close agreement with the actual values. In summary, this study highlights the NN model as a robust choice for forecasting solar energy production. This means the model can produce reliable data analysis or prediction task results. Additionally, the prediction results of the NN algorithm were compared with actual data. This comparison examined the statistical properties of the predicted and actual data and generally found that the predictions were quite close to the actual values. However, there are significant differences in some minimum and maximum values. Finally, in a section where the multidimensional scaling (MDS) method was used, the data distribution estimated by the NN algorithm was visualized. This visualization shows that the forecast data is partially clustered, but the clustering classes are limited.
Finally, this study demonstrates that the use of the NN algorithm in solar energy predictions results in more precise forecasts of solar energy production, as it effectively captures the intricacies and interplay of environmental factors. This algorithm has provided high accuracy in solar energy production forecasts, especially its ability to process large data sets and recognize complex patterns. In addition to explaining the relationship between solar energy production and environmental factors, this study will provide insight into the potential of increasing the efficiency of solar energy facilities thanks to the accuracy of the predictions obtained with the NN algorithm.

Author Contribution

Yasemin Ayaz Atalan contributed to all stages of the study.

Conflict of Interest

The author has declared no conflict of interest.

REFERENCES

[1] Y.A. Atalan and A. Atalan, “Integration of the Machine Learning Algorithms and I-MR Statistical Process Control for Solar Energy,” Sustainability, vol. 15, no. 18, p. 13782, Sep. 2023.

[2] N. Fatima, Y. Li, M. Ahmad, G. Jabeen, and X. Li, “Factors influencing renewable energy generation development: a way to environmental sustainability,” Environ. Sci. Pollut. Res., vol. 28, no. 37, pp. 51714–51732, 2021.

[3] V. Mhasawade, Y. Zhao, and R. Chunara, “Machine learning and algorithmic fairness in public and population health,” Nat. Mach. Intell., vol. 3, no. 8, pp. 659–666, Aug. 2021.

[4] V. Ramanathan and Y. Feng, “Air pollution, greenhouse gases and climate change: Global and regional perspectives,” Atmos. Environ., vol. 43, no. 1, pp. 37–50, 2009.

[5] L. Qi and Y. Zhang, “Effects of solar photovoltaic technology on the environment in China,” Environ. Sci. Pollut. Res., vol. 24, pp. 22133–22142, 2017.

[6] A. Sharif, S.A. Raza, I. Ozturk, and S. Afshan, "The dynamic relationship of renewable and nonrenewable energy consumption with carbon emission: a global study with the application of heterogeneous panel estimations," Renew. energy, vol. 133, pp. 685–691, 2019.

[7] F. Dincer, "The analysis on photovoltaic electricity generation status, potential and policies of the leading countries in solar energy," Renew. Sustain. energy Rev., vol. 15, no. 1, pp. 713–720, 2011.

[8] N. Armaroli and V. Balzani, "The future of energy supply: challenges and opportunities," Angew. Chemie Int. Ed., vol. 46, no. 1-2, pp. 52–66, 2007.

[9] S. Kanwal, M.T. Mehran, M. Hassan, M. Anwar, S. R. Naqvi, and A.H. Khoja, "An integrated future approach for the energy security of Pakistan: Replacement of fossil fuels with syngas for better environment and socio-economic development," Renew. Sustain. Energy Rev., vol. 156, p. 111978, 2022.

[10] U. Pelay, L. Luo, Y. Fan, D. Stitou, and M. Rood, "Thermal energy storage systems for concentrated solar power plants," Renew. Sustain. Energy Rev., vol. 79, pp. 82–100, 2017.

[11] F. Trieb and H. Elnokraschy, "Concentrating solar power for seawater desalination," IWCT, vol. 12, pp. 2–13, 2007.

[12] A.C. Şerban and M.D. Lytras, "Artificial intelligence for smart renewable energy sector in europe—smart energy infrastructures for next generation smart cities," IEEE access, vol. 8, pp. 77364–77377, 2020.

[13] A. Mohammad and F. Mahjabeen, "Revolutionizing Solar Energy: The Impact of Artificial Intelligence on Photovoltaic Systems," Int. J. Multidiscip. Sci. Arts, vol. 2, no. 1, 2023.

[14] A. Gopi, P. Sharma, K. Sudhakar, W. K. Ngui, I. Kirpichnikova, and E. Cuce, "Weather impact on solar farm performance: A comparative analysis of machine learning techniques," Sustainability, vol. 15, no. 1, p. 439, 2022.

[15] M.M.V. Cantarero, "Of renewable energy, energy democracy, and sustainable development: A roadmap to accelerate the energy transition in developing countries," Energy Res. Soc. Sci., vol. 70, p. 101716, 2020.

[16] 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.

[17] A. Atalan, H. Şahin, and Y.A. Atalan, "Integration of Machine Learning Algorithms and Discrete-Event Simulation for the Cost of Healthcare Resources," Healthcare, vol. 10, no. 10, p. 1920, Sep. 2022.

[18] F. Rodríguez, A. Fleetwood, A. Galarza, and L. Fontán, "Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control," Renew. energy, vol. 126, pp. 855–864, 2018.

[19] 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," Renew. Sustain. Energy Rev., vol. 124, p. 109792, 2020.

[20] O. Erdinc and M. Uzunoglu, "Optimum design of hybrid renewable energy systems: Overview of different approaches," Renew. Sustain. Energy Rev., vol. 16, no. 3, pp. 1412–1425, 2012.

[21] Z. He, W. Guo, and P. Zhang, "Performance prediction, optimal design and operational control of thermal energy storage using artificial intelligence methods," Renew. Sustain. Energy Rev., vol. 156, p. 111977, 2022.

[22] I.M. Galván, J.M. Valls, A. Cervantes, and R. Aler, "Multi-objective evolutionary optimization of prediction intervals for solar energy forecasting with neural networks," Inf. Sci. (Ny)., vol. 418–419, pp. 363–382, 2017.

[23] A. Kuzmiakova, G. Colas, and A. McKeehan, "Predicting solar energy using machine learning: CS 229 project," 2023. https://github.com/adelekuzmiakova/CS229machine-learning-solar-energy-predictions/tree/master

[24] H. İnaç, Y.E. Ayözen, A. Atalan, and C.Ç. Dönmez, "Estimation of Postal Service Delivery Time and Energy Cost with E-Scooter by Machine Learning Algorithms," Appl. Sci., vol. 12, no. 23, p. 12266, Nov. 2022.

[25] J. Li, M.S. Herdem, J. Nathwani, and J.Z. Wen, "Methods and applications for artificial intelligence, big data, internet-of-things, and blockchain in smart energy management," Energy AI, p. 100208, 2022.

[26] S. Osama, H. Shaban, and A.A. Ali, "Gene reduction and machine learning algorithms for cancer classification based on microarray gene expression data: A comprehensive review," Expert Syst. Appl., vol. 213, p. 118946, 2023.

[27] A. Abraham et al., "Machine learning for neuroimaging with scikit-learn," Front. Neuroinform., vol. 8, p. 14, 2014.

[28] K. Hsu, S. Levine, and C. Finn, "Unsupervised learning via meta-learning," arXiv Prepr. arXiv1810.02334, 2018.

[29] R. Diao, Z. Wang, D. Shi, Q. Chang, J. Duan, and X. Zhang, "Autonomous voltage control for grid operation using deep reinforcement learning," in 2019 IEEE Power & Energy Society General Meeting (PESGM), IEEE, 2019, pp. 1–5.

[30] J. Yang et al., "Neuromorphic engineering: from biological to spike-based hardware nervous systems," Adv. Mater., vol. 32, no. 52, p. 2003610, 2020.

[31] S. Boccaletti et al., "The structure and dynamics of multilayer networks," Phys. Rep., vol. 544, no. 1, pp. 1–122, 2014.

[32] M. Khandelwal and T.N. Singh, "Prediction of blast-induced ground vibration using artificial neural network," Int. J. Rock Mech. Min. Sci., vol. 46, no. 7, pp. 1214–1222, 2009.

[33] D. Graupe, Principles of Artificial Neural Networks, vol. 7. in Advanced Series in Circuits and Systems, vol. 7. World Scıentıfıc, 2013.

[34] C. Coman, L.G. Țîru, L. Meseșan-Schmitz, C. Stanciu, and M.C. Bularca, "Online teaching and learning in higher education during the coronavirus pandemic: Students' perspective," Sustainability, vol. 12, no. 24, p. 10367, 2020.

[35] V. Nasteski, "An overview of the supervised machine learning methods," Horizons. b, vol. 4, pp. 51–62, 2017.

[36] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," PLoS One, vol. 13, no. 3, p. e0194889, 2018.

[37] Y.-C. Wang and J.M. Usher, "Application of reinforcement learning for agent-based production scheduling," Eng. Appl. Artif. Intell., vol. 18, no. 1, pp. 73–82, 2005.

[38] E.L. Jacobsen and J. Teizer, "Deep learning in construction: Review of applications and potential avenues," J. Comput. Civ. Eng., vol. 36, no. 2, p. 3121001, 2022.

[39] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," J. Ind. Inf. Integr., vol. 23, p. 100224, 2021.

[40] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," J. big Data, vol. 8, pp. 1–74, 2021.

[41] D.P. Kiley, S. Haley, B. Saylor, and B.L. Saylor, "The Value of Evidence-Based Computer Simulation of Oral Health Outcomes for Management Analysis of the Alaska Dental Health Aide Program," Institute of Social and Economic Research, University of Alaska Anchorage, 2008. [Online]. Available: http://hdl.handle.net/11122/4459

[42] M.S. Lewis-Beck and A. Skalaban, "The R-squared: Some straight talk," Polit. Anal., vol. 2, pp. 153–171, 1990.

[43] A. Atalan, "Statistical optimization of forecast data from Adaptive Boosting and Support Vector Machine Algorithms," in International Conference on Engineering, Natural and Social Sciences, 2023, pp. 571–579.

[44] C.J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Clim. Res., vol. 30, no. 1, pp. 79–82, 2005.

[45] 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 Comput. Sci., vol. 7, p. e623, 2021.

[46] A. Atalan, "Effect of healthcare expendıture on the correlatıon between the number of nurses and doctors employed," ınt. J. Heal. Manag. Tour., vol. 6, no. 2, pp. 515–525, 2021.

[47] D.B. Figueiredo Filho, J. A. S. Júnior, and E. C. Rocha, "What is R2 all about?," Leviathan (São Paulo), no. 3, pp. 60–68, 2011.

[48] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), IEEE, Oct. 2011, pp. 528–533.

[49] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks," in 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Oct. 2016, pp. 002858–002865.

[50] Z. Li, S. Rahman, R. Vega, and B. Dong, "A Hierarchical Approach Using Machine Learning Methods in Solar Photovoltaic Energy Production Forecasting," Energies, vol. 9, no. 1, p. 55, 2016.