Kütahya Dumlupınar University
Institute of Graduate Studies

Journal of Scientific Reports-A
E-ISSN: 2687-6167
Number 51, December 2022

RESEARCH ARTICLE

SOLAR RADIATION FORECAST by USING MACHINE LEARNING METHOD for GAZIANTEP PROVINCE

Yusuf Alper KAPLAN[1], Emre BATUR[2], Gülizar Gizem TOLUN[3,*]

[1]Osmaniye Korkut Ata University, Faculty of Engineering, Energy Systems Engineering, Osmaniye, alperkaplan@osmaniye.edu.tr, ORCID: 0000-0003-1067-110X
[2]Osmaniye Korkut Ata University, Faculty of Engineering, Energy Systems Engineering, Osmaniye, emrebatur.esm@gmail.com, ORCID: 0000-0002-7538-7575
[3*]Osmaniye Korkut Ata University, Faculty of Engineering, Energy Systems Engineering, Osmaniye, gulizargizemunaldi@osmaniye.edu.tr, ORCID: 0000-0003-1876-9283

Receive Date:10.09.2022 *Accepted Date: 07.10.2022*

ABSTRACT

Renewable energy sources have become a popular topic all over the world in terms of cost, efficiency and environmental pollution. Solar energy is the most significant of the renewable energy sources. Solar energy, which was used only as heat and light energy in the past, is widely used in electrical energy production with the advancement of today's technology. Traditionally used photovoltaic cells are semiconductor materials that are produced in various chemical structures and convert the energy they receive from sunlight directly into electrical energy. The research and development of photovoltaic cells is moving forward at an accelerating pace. With this development process and relying on the today's technology, it is aimed to increase the efficiency of photovoltaic cells and to produce more electrical energy as a result of various trials. By analysing the energy production of photovoltaic cells, efficiency-enhancing situations are examined according to solar radiation values. In this study, a model was constructed using the regression approach, which is a method of machine learning. This model has been developed using the MATLAB program of the meteorological data of 2021 from Gaziantep. In addition, a variety of error analysis tests were utilized in order to evaluate the effectiveness of the model that was built. As a consequence, the model created using the linear regression method yields successful results in estimating solar radiation in Gaziantep province. This is demonstrated by the coefficient of determination (R^2) value of 0.98, the Mean Absolute Error (MAE) value of 0.023, the Root Mean Square Error (RMSE) value of 0.028, and the Mean Square Error (MSE) value of 0.0008.

Keywords: *Solar energy, Energy production, Photovoltaic cell, Cell efficiency, Machine learning*

1. INTRODUCTION

Depending on the development and progress of today's technology, energy consumption is increasing in all areas of life. In the current world order, energy needs are fulfilled with fossil-based resources such as coal, natural gas, and oil. However, fossil fuels are not energy sources that can be used in the long term due to their irreversible damage to the environment and their tendency to run out [1].

The intensity of solar energy before reaching the earth's atmosphere is approximately 1370 W/m², but the amount of solar energy reaching the earth's surface varies between 0-1100 W/m² due to the atmosphere [2]. A large part of this energy is used by plants. The remaining energy is converted into heat and is currently used in the formation of many atmospheric events. More specifically, Turkey has a high potential in terms of the efficiency level that can be obtained from solar energy. Figure 1 shows the regional solar radiation map of Turkey.

Figure 1. Total Solar Radiation Map of Turkey [3].

According to Turkey's solar energy potential map, the annual total sunshine duration is 2741 hours/year, and the annual total incoming solar energy is 1527.46 kWh/m^2-year. According to Table 1, the average daily sunshine duration in Turkey is 7.5 hours [4].

Table 1. Solar radiation and sunshine duration values of Gaziantep province.

Months	**Total Solar Energy per Day** (kWh/m^2-day)	**Total Solar Energy per Month** $(kWh/m^2-month)$	**Sunshine Duration per Day** $(hours/day)$	**Sunshine Duration per Month** $(hours/month)$
January	1,79	55,49	4,11	127,41
February	2,5	70	5,22	146,16
March	3,87	119,27	6,27	194,37
April	4,93	147,9	7,46	223,8
May	6,14	190,34	9,1	282,1
June	6,57	197,1	10,81	324,3
July	6,5	201,5	11,31	350,61
August	5,81	180,11	10,7	331,7
September	4,81	144,3	9,23	276,9
October	3,46	107,26	6,87	212,97

November	2,14	64,2	5,15	154,5
December	1,59	49,29	3,75	116,25
Total	50,11	1527,46	89,98	2741,07
Average	4,18	127,29	7,50	228,42

Photovoltaic panels, which are used in the production of electricity from solar energy, provide production at different powers at different points of our country. Regional, seasonal, and environmental factors are at the forefront of these differences. This study forms the basis of examining the production of photovoltaic cells in different provinces and estimating the energy production in these provinces using a machine learning method called linear regression. Based on this estimation process, investors and institutions could be able to determine the feasibility of benefiting from solar energy, predict the amortization period, maximize energy production efficiency, and determine the best region to invest in. In the estimation part, the solar radiation data of Gaziantep province and the sunshine duration parameters were processed in the MATLAB environment and carried out by the linear regression method. In the MATLAB environment, the obtained data were estimated using the linear regression method. By means of these method, production values and error values were determined before the solar power plant was installed.

Kaplan used the moving least squares approach (MLSA) and the least squares method (LSM) to calculate the coefficients of the Weibull distribution function. In this study, the outcomes of the MLSA graph approach were compared against the results of the least squares method on a monthly and annual basis [5]. Atique et al. used auto-regressive integrated moving average (ARIMA) to model the time series with its seasonal variant and then compared the performance with support vector machine (SVM) and artificial neural network (ANN). The analysis revealed that SVM performed best, however it was proposed that the reasons behind ANN's low efficiency should be investigated [6]. Aksoy and Selbaş utilized machine learning methods to estimate energy values based on wind speed, temperature, and direction data acquired from a wind turbine. A mathematical equation was proposed by researchers who achieved 90% accuracy in their studies [7]. Abdelhafidi et al. developed a model based on global solar radiation estimation using temperature, cloud cover, insolation rate, evaporation, wind speed, sunshine duration, visibility, and global solar radiation data from 1986–2014 obtained from the Oran radiometric station in Algeria. A stepwise multiple linear regression analysis was performed to fit the meteorological variables to the global solar radiation data [8]. Nath et al. used two different machine learning approaches to estimate solar energy. In their study, they used Neural Network (NN) and long short-term memory (LSTM) and determined that NN performed better [9].

2. CALCULATION of SOLAR RADIATION

While calculating the solar radiation required for the energy production of photovoltaic systems, the monthly average solar radiation is used. The development of the Angstrom model was inspired by the meteorological, regional, astronomical, geographical, and geometric parameters of global solar radiation [10].

JSR A
Journal of Scientific Reports

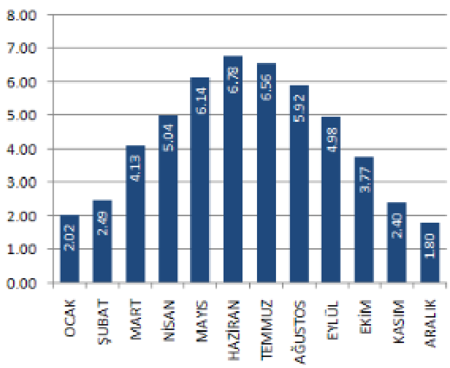
$$\frac{H}{H_0} = a + b\left(\frac{S}{S_0}\right) \tag{1}$$

When this equation is examined in detail, H represents the monthly average daily global solar radiation of the determined region. S is the average daily sunshine duration of the monthly determined region, H_0 is the daily solar radiation value of the horizontal surface of the region, and S_0 is the astronomical day length. The specified a and b coefficients are the regression constants to be determined. The extraterrestrial solar radiation coming to the horizontal surface of the determined region is calculated by Equation 2 [11].

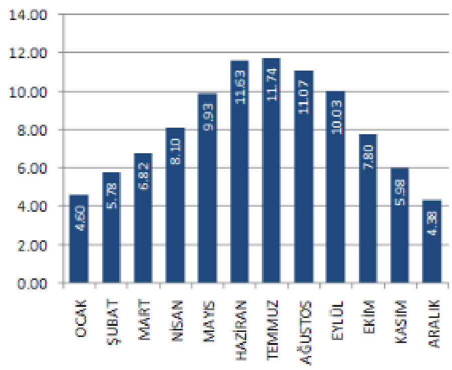
$$H_0 = \frac{24}{\pi} k\left[\cos(\phi)\cos(\delta)\sin(\omega_s) + \left(\frac{\pi}{180}\right)\sin(\phi)\sin(\delta)\omega_s\right] \tag{2}$$

In this equation, the value of k is the day of the year. δ is the sun declination angle, Ø is the latitude angle of the region, and ωs is the average sunrise value for a certain month according to the angles of the region. Although Isc is expressed as solar constant, it is accepted as 1367 W/m^2 [11].

$$\omega_s = \cos^{-1}\left[-\tan(\delta)\tan(\phi)\right] \tag{3}$$



Gaziantep Global Solar Radiation Values
$(kWh/m^2 - day)$



Gaziantep Sunshine Duration
$(hours)$

Figure 2. Solar Radiation and Sunshine Duration of 2021 for Gaziantep Province [3].

Figure 2 illustrates the global solar radiation and sunshine duration of Gaziantep province, which is the region determined in this study.

Table 2. Solar radiation and sunshine duration values of Gaziantep province.

Months	H	H_0	H/H_0	S	S_0	S/S_0
January	2020,000	4742,326	0,426	4,600	9,772	0,471
February	2490,000	6103,072	0,408	5,780	10,626	0,544
March	4130,000	7993,983	0,517	6,820	11,759	0,580
April	5040,000	9841,157	0,512	8,100	12,967	0,625
May	6140,000	11084,258	0,554	9,930	13,987	0,710
June	6780,000	11570,674	0,586	11,630	14,503	0,802
July	6560,000	11301,513	0,580	11,740	14,259	0,823
August	5920,000	10274,358	0,576	11,070	13,371	0,828
September	4980,000	8606,105	0,579	10,030	12,201	0,822
October	3770,000	6650,105	0,567	7,800	10,996	0,709
November	2400,000	5037,583	0,476	5,980	9,984	0,599
December	1800,000	4324,769	0,416	4,380	9,495	0,461

S_0 represents the astronomical day length obtained from the movement of the earth around the sun [11].

$$S_0 = \frac{2}{15}\omega_s \tag{4}$$

The sun declination angle δ is the angle that the sun rays coming to the earth make with the equatorial plane. This angle varies between -23.45˚ and +23.45˚ during the year. The value of n in the formula is the number of days of a year starting from January [12].

$$\delta = 23.45\sin\left(360\frac{n+284}{365}\right) \tag{5}$$

3. ESTIMATION PROCESS for GAZIANTEP PROVINCE

Gaziantep, which is located at the junction of the Mediterranean Region and the South-eastern Anatolia Region, is located between 36° 28' and 38° 01' east longitudes and 36° 38' and 37° 32' north latitudes [13]. Average global solar radiation H and mean sunshine duration S are shown in Figure 2. By obtaining H_0 and S_0 values, a and b coefficients were obtained by using the machine learning method.

The above-mentioned average solar radiation H, extra-atmospheric solar radiation H_0, average sunshine duration S, astronomical day length S_0, and the values of the H/H_0 and S/S_0 to be used when constructing the Angstrom model are given in Table 2. The datasets in Table 2 were processed in the MATLAB program, and the regression graphs and values were obtained by processing the 2021 average solar radiation data and sunshine duration data of Gaziantep province. The model was found by using the linear regression machine learning method over the MATLAB by using the average solar radiation value and average sunshine duration for Gaziantep province. In the model found, error analysis methods such as RMSE, MAE, MSE, and R^2 were applied, and the accuracy of the model was revealed as shown in Table 3.

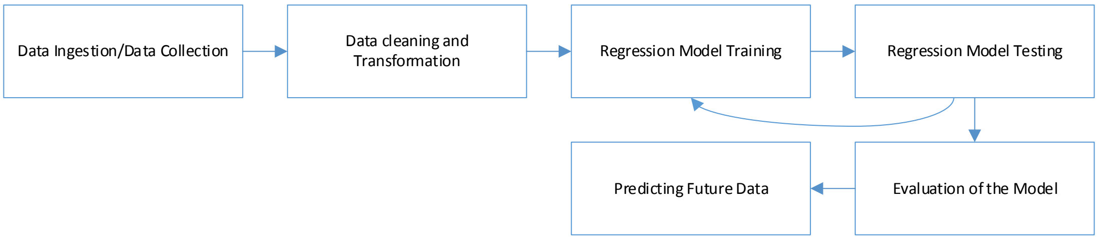


Figure 3. Flowchart of the Regression Process [14].

In this study, the linear regression model was utilized to forecast future values. Figure 3 illustrates a flow chart of the complete operation. The first data for the learning process is collected and processed in the data storage or data ingestion section. The dataset was collected from the Solar Potential Atlas (GEPA) [3]. The converted data is trained and tested on the Linear Regression Model in the Regression Model Training and Regression Model Testing sections. The training and testing process is repeated numerous times by modifying settings until the best accuracy is reached. In the evaluation model, regression metrics are measured. Future values can be predicted using the best model with satisfied metrics [14].

JSR A
Journal of Scientific Reports

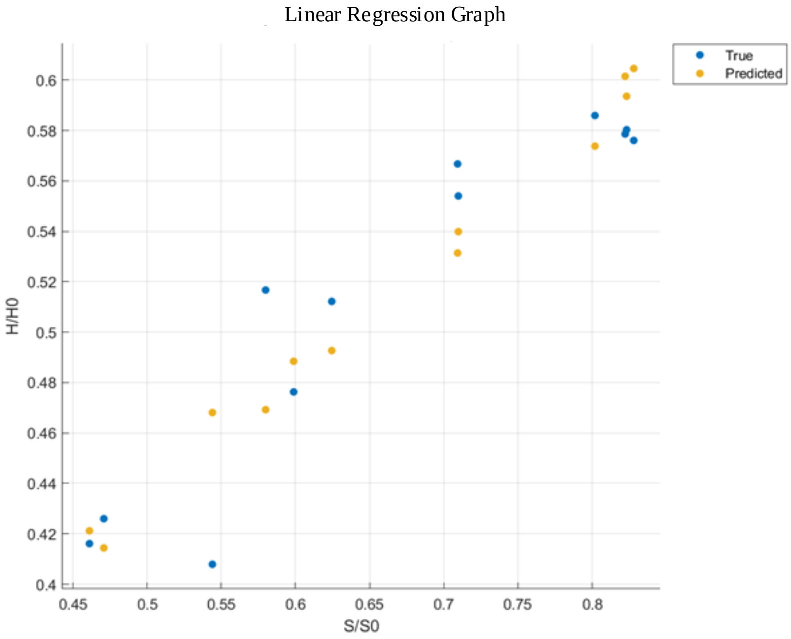


Figure 4. Linear Regression Graph.

Figure 4 shows the comparison of actual and estimated data on the graph. The model created is based on the data in this graph.

$$\frac{H}{H_0} = -0{,}1898 + 0{,}4918\left(\frac{S}{S_0}\right) \tag{6}$$

The model in Equation 6 is obtained to estimate the solar radiation for selected region. The linear model that emerges as a result of the linear regression method.

Table 3. Regression model and error parameters.

Regression Model	a	b	Model	MAE	RMSE	R^2	MSE
Linear Regression	-0,1898	0,4918	$y = 0{,}4918x - 0{,}1898$	0,023	0,028	0,98	0,0008

Four distinct statistical tests were used to evaluate the performance of the model constructed for the Gaziantep region. As demonstrated in Table 3, the created model performed accurately in terms of estimating solar radiation.

4. RESULTS and DISCUSSION

The increase in the need for energy used in the world day by day proves that the current energy is not at a sufficient level. Today, the increase in the cost of fossil fuels, which are the most widely used to satisfy the world's energy needs, and the irreversible damage to the environment have led humanity to seek alternative sources. Renewable energy sources have emerged as a result of this search, and solar energy is seen as the most popular among these sources.

The utilization of developing technology has increased due to the development of computer science and the emergence of new methods. The use of artificial intelligence, machine learning methods, and many different algorithms paves the way for the solar energy sector. In this study, it is aimed to create a new model by using linear regression from machine learning methods in the MATLAB program. Machine learning methods are one of the most efficient methods of estimation.

Machine learning algorithms have grown in popularity for estimating solar radiation. The linear regression method was chosen from among numerous machine learning algorithms due to its ease of use. However, different machine learning methods can also be used to predict solar radiation in the selected region for a higher accuracy and more precise estimation process. When the estimated experimental results analysis for Gaziantep province was examined, the results were obtained according to the MATLAB program, with the linear regression results as RMSE is 0.028, MAE is 0.023, MSE is 0.0008, and the coefficient of determination R^2 is evaluated as 0.98. When the results were analysed, it was seen that realistic results could be obtained when the linear regression method was tried for Gaziantep province. By comparing the similar studies in the literature, it is clearly observed that the developed model gives accurate results. In a conclusion, it is seen that each province-based region can be modelled by using the machine learning method in terms of solar radiation intensity, and sunshine duration. The linear regression model could be utilized in order to estimate solar radiation for various regions.

ACKNOWLEDGEMENT

The author declares that there are no conflict of interests.

REFERENCES

[1] Sayin, S., Koc, İ., (2011), Photovoltaic (PV) Systems used in active utilization of solar energy and usage methods in building, Selcuk University Journal of Engineering, Science and Technology, 26, 89-106.

[2] Sahan, M., Sahan, H., and Yegingil, I., (2014), Measuring annual total and ultraviolet (UV) solar energy data, Journal of Suleyman Demirel University Graduate School of Natural and Applied Sciences, 14, 10–16.

[3] (2022), Solar Energy Potential Atlas. [Online]. Available: https://gepa.enerji.gov.tr/Mycalculator/Default.Aspx

[4] Canka Kilic, F., (2015), Latest status in Turkey and production technologies, Engineer and Machine, 56, 28-40.

[5] Kaplan, A.G., (2022), A new approach based on moving least square method for calculating the Weibull coefficients, Environmental Progress & Sustainable Energy, 41(4), e13934.

[6] Atique, S., Noureen, S., Roy, V., and Macfie, J., (2020), Time series forecasting of total daily solar energy generation: A comparative analysis between ARIMA and machine learning techniques, 2020 IEEE Green Technologies Conference, 1-3 April 2020, 175-180.

[7] Aksoy, B., and Selbas, R., (2019), Estimation of wind turbine energy production value by using machine learning algorithms and development of implementation program, Energy Sources Part A: Recovery, Utilization and Environmental Effects, 43, 692-704.

[8] Abdelhafidi, N., Bachari, N.I., and Abdelhafidi, Z., (2021), Estimation of solar radiation using stepwise multiple linear regression with principal component analysis in Algeria, Meteorology and Atmospheric Physics, 133, 205-216.

[9] Nath, N.C., Sae-Tang, W., and Pirak, C., (2020), Machine learning-based solar power energy forecasting, Journal of the Society of Automotive Engineers Malaysia, 4, 307-322.

[10] [Besharat, F., Dehghan, A.A., and Faghih, A.R., (2013), Empirical models for estimating global solar radiation: A review and case study, Renewable and Sustainable Energy Reviews, 21, 798-821.

[11] Kaplan, Y.A., Sarac, M.S., and Unaldi, G.G., (2022), Developing a new model in solar radiation estimation with genetic algorithm method, Environmental Progress & Sustainable Energy.

[12] Sahan, M., (2021), Global solar radiation forecast for Gaziantep, Antakya and Kahramanmaras using artificial neural networks and Angström-Prescott equations, Suleyman Demirel University Faculty of Arts and Science Journal of Science, 16, 368-384.

[13] Roger, A., and Ventre, M.J., (2003), Photovoltaic Systems Engineering, 2nd ed., CRC Press.

[14] Killada, P., (2017), Data analytics using regression models for health insurance market place data, MSc. Thesis in Engineering: Computer Science, The University of Toledo.