Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi
International Journal of Engineering Research and Development

UMAGD, (2023) 15(2), 704-713.

10.29137/umagd.1268055

Volume:15 Sayı/Issue:2 June 2023

Research Article

Forecasting Solar Radiation Based on Meteorological Data Using Machine Learning Techniques: A Case Study of Isparta Province

Buğra Güzel[1] *, Onur Sevli[2], Ersan Okatan[1]

[1] Mehmet Akif Ersoy University, Gölhisar School of Applied Sciences, Department of Computer Technologies and Information Systems, 15400 Burdur, TÜRKİYE
[2] Mehmet Akif Ersoy University, Faculty of Engineering-Architecture, Department of Computer Engineering, 15030, Burdur, TÜRKİYE

***Received**: 21/03/2023* ***Accepted**: 06/06/2023* ***Published Online:** 30/06/2023*
***Final Version**: 30/06/2023*

Abstract
Solar energy systems which is one of the renewable energy sources takes more interest and gains prevalence day by day. A significant problem in solar energy systems as in other many renewable energy sources is the instability of the energy that the system will provide. Forecasting of the energy to be obtained is very important in this respect. In this study, solar radiation has been forecasted using meteorological data taken from the General Directorate of Meteorology for Isparta province. Random Forest (RF), k-Nearest Neighbor (k-NN), Artificial Neural Network (ANN) and Deep Learning (DL) methods have been used for forecasting. In addition, the results of dummy variable usage for time data have been examined with these different methods. According to the findings obtained, it is seen that the dummy variable usage increases performance for ANN and DL methods but decreases performance for RF and k-NN methods. Best results have been obtained with ANN and DL for the forecasting of the solar radiation.

Key Words
"Artificial neural networks, deep learning, dummy variable, random forest, solar radiation forecast"

***Corresponding Author:** bguzel@mehmetakif.edu.tr*

1. Introduction

Due to economic and environmental reasons, more and more investments are made in renewable energy systems and the interest is increasing (Al-Sbou and Alawasa, 2017). According to the report prepared by the International Renewable Energy Agency (IRENA), the energy capacity obtained from renewable energy sources reached 2799 GW at the end of 2020, 26% of which was solar energy systems. In 2020, there was an increase of 10.3% in renewable energy capacity all over the world, and 91% of this increase was provided by solar and wind energy (Anonymus, 2021). This data shows that solar and wind energy will rapidly exceed the rate of other systems in renewable energy systems. The fact that, unstable and discontinuous nature of the energy obtained from renewable energy systems is a disadvantage and causes significant technical problems. Forecasting the energy to be produced makes an important contribution to overcoming these technical problems like sustainability, stability and security of the energy source (Al-Sbou and Alawasa, 2017).

Several studies were made to forecast solar energy produced. In the study of Al-Rousan et al. (2021), the dataset was collected by using a designed sun tracking system. They examined the ML algorithms in 5 categories, and obtained the best results with M5rules, RF, Bagging, k-NN and Multilayer Perceptron (MLP) algorithms in each category. Among these algorithms, the best performance result was obtained with RF. Ozoegwu (2019) proposed an alternative hybrid ANN model to the Nonlinear Autoregressive (NAR) based ANN method in his study and obtained better results. Elsheikh et al. (2019) worked on the modeling of different systems using solar energy with ANN and examined different ANN models. Alomari et al. (2018) made an hourly solar power forecast one day ahead with the solar radiation and solar energy power values taken previously in their study. They obtained results with an R value of 0.99. Ayko and Bozkurt Keser (2021), on the other hand, compared 45 different ML algorithms in their study. They found that ensemble learning methods outperformed other ML algorithms. Erten and Aydilek (2022) compared 4 different regression models as Lasso, Ridge, Elastic and Linear in solar power prediction. In their study, Elastic Regression model outperformed other proposed models.

There is also a time series approach with Long Short-Term Memory Network in the literature for solar radiation forecast (Kara, 2019). Sanders et al. (2017) used weather forecast values together with actual weather data. In the results obtained, they reached 40.2% better average absolute error values in 24-hour forecast values and 7.6% better in 1-hour forecast values. Hamdan et al. (2017) reached Squared Correlation (R^2) values of 97% with Feed Forward and Elman, and 99% with Nonlinear Autoregressive Exogenous (NARX) from ANN models in order to forecast hourly solar radiation values. Ibrahim and Khatib (2017) used Firefly Algorithm (FFA) and RF algorithms together in their studies and they determined the ideal number of trees in the RF with FFA firstly. With this hybrid model they concluded that better performance results could be obtained than ANN, ANN-FFA and RF methods. In their study, Al-Sbou and Alawasa (2017) used hourly weather data to forecast solar radiation 24 hours in advance with the NARX method, and achieved the best performance by using all three of the temperature, wind and humidity values. Yadav et al. (2015) used meteorological data taken from 76 different points in their studies with different ANN methods and they concluded that the variables of temperature, altitude and sun hour are the most effective, while the variables of aperture index, extra-atmospheric solar radiation, latitude and longitude are less effective. Mohammed et al. (2013) used hourly meteorological data and reached the best results with the Marquardt–Levenberg algorithm made with the NARX model in their forecast. Demirtas et al. (2012) concluded that the forecasting results obtained by using k-NN are more successful than other methods such as ANN and Fuzzy in their study.

In his study, Şahin (2013) used month, latitude, longitude and surface temperature data and made forecasting with different ANN models for solar radiation. In another study, Şeker (2021) obtained high-accuracy values by using ANN and concluded that ANN can be used for solar radiation forecast. Şeker (2021) used minute, hour, day and month variables as independent variables in the ANN model. In addition, there are models that forecast solar power generation in the literature (Uğuz et al., 2019).

The use of other coding techniques for variables that do not have numerical superiority but are expressed numerically by nature may affect the performance in ML. For example; although the year expressions do not represent a numerical superiority, they are mentioned numerically (such as 2022, 2023). There are occasional contradictions in the use of variables in the models to be built with such data. Although there are various methods for coding such data, the most used method is the use of dummy variables (Jolly and Gupta, 2021). There are examples in the literature where time variables are used as dummy variables (Hong et al., 2010; Pardo et al.,2002).

The aim of this study is to create a successful 1-hour ahead solar radiation forecasting model for the Isparta province by using different ML algorithms. In addition, the effects of the use of time variables as numerical or dummy on the performance of the model was examined. Time variables (year, month, day and hour) were used as independent variables together with meteorological data.

This article is structured as follows. We introduce the dataset and brief theory of ML models in Section 2. In Section 3, we described the obtained results and discussion. Finally, some concluding remarks are given in Section 4.

2. Material and Methods

In recent years, thanks to numerous algorithms and increasing processor capacities, ML has proven its forecasting power in data science and has gained an important place in the big data world. It has been widely used in studies on forecasting solar radiation and continues to be used with new techniques. But the performance of the algorithm depends on the structure of the dataset; such as feature types, sample size. Models should be tested with the existing dataset to select the best algorithm.

In this section, analysis of the data set and ML methods used were given. The hourly total global solar radiation, wind speed, temperature, relative humidity and cloudiness data of Isparta province were received. The data set was compiled by putting these meteorological data and also year, month, day and time values into separate variables.

RF, k-NN, ANN and DL methods were tested in order to forecast the output of solar radiation with the input of hour, day, month, year, temperature, cloudiness, relative humidity and wind speed. Rapidminer ML software platform has been used to create and test all models. In addition to testing four different ML methods, the effect of using time variables (year, month, day, hour) as numerical or categorical on the model performance was examined. For this purpose, all the used integer time variables were converted into dummy variables and the models were reconstructed and the results have been compared.

2.1. Dataset

The hourly meteorological data covering the 2016- 2020 of Isparta province, which were taken separately from the General Directorate of Meteorology, were brought together with the help of unique codes created with time characteristics. The created data set contains 9 features including year, month, day, hour, temperature, cloudiness, relative humidity and wind speed as inputs and solar radiation as output. Also, the dataset has 40557 samples. The lowest, highest, average and standard deviation values of the meteorological features of the created data set are given in Table 1. Additionally, the correlation matrix of the meteorological features used is given in Table 2.

Table 1. Statistical Values of Dataset

Features	Minimum	Maximum	Mean	Standard Deviation
Solar Radiation (W/m2)	0	61251	11496.37	16800.92
Temperature (^{0}C)	-13.70	37.40	13.26	9.46
Cloudiness (Okta)	0	8	2.58	2.65
Relative Humidity (%)	7	99	60.93	22.10
Wind Speed (m/s)	0	20.90	1.67	1.27

Table 2. Correlation Matrix for Meteorological Data

Features	Solar Radiation	Temperature	Wind Speed	Relative Humidity	Cloudiness
Solar Radiation	1.00				
Temperature	0.50	1.00			
Wind Speed	0.29	0.19	1.00		
Relative Humidity	-0.56	-0.73	-0.34	1.00	
Cloudiness	-0.10	-0.19	0.30	0.28	1.00

2.2. Machine Learning Algorithms

Solar Radiation forecast could be made by using 4 different methods, which are physical, empirical, statistical and ML. Here, ML methods have been used because these models have higher accuracy and wider application compared to others (Zhou et al., 2021). In the study the total hourly global solar radiation value has been forecasted by using most powerful machine learning methods which are RF, k-NN, ANN and DL. These methods are briefly described in below.

2.2.1. Random Forest (RF)

Decision Tree (DT) is a ML algorithm that aims to forecast by defining rules in a certain order towards nodes or leaves. RF is a method created by bringing together more than one DT to produce more consistent results. In the RF method, the trees in the community are formed randomly, so less correlation occurs between the trees and the forecasting accuracy is higher (Korkmaz et al., 2018).

In some studies, the RF algorithm is used for the forecast of solar radiation. The RF has good accuracy compared to other ML techniques in most of the studies (Bamisile et al., 2022; Benali et al., 2019; Torres-Barrán et al., 2019; Zeng et al., 2020). Using the RF, interactions between variables and feature selection naturally occurs while learning process. Also, variable importance measures and taking the ordering information into account are extreme features of RF over other ML methods (Sun et al., 2016). In addition, the RF has a few parameters and does not overfit (Ibrahim and Khatib, 2017). For this study the least square method was used and number of trees was 100 and also maximum tree node depth was used as 10.

2.2.2. K Nearest Neighbor (k-NN)
The k-NN algorithm is a ML algorithm that can be used in both regression and classification forecasting problems (Demolli et al., 2019). The k-NN algorithm have been rarely used in energy systems area (Marzouq et al., 2019; Yesilbudak and Ozcan, 2022). In this study, the k-NN is used because of it is power on forecast and for the purpose of comparison. The effect of using dummy variable on the k-NN model was examined and results are analyzed. For the k-NN model, k used as 5 and Euclidean distance measurement used to determine the nearest neighbors.

2.2.3. Artificial Neural Network (ANN)
The billions of nerve cells (neurons) in the human brain try to produce solutions to complex problems in interaction with each other, using the information they have previously learned. Neurons in an ANN algorithm are similarly imitating this, learning the input data in general terms and trying to produce consistent solutions to problems (Demirtas et al., 2012).

The neuron structure in ANN is mainly as in Figure 1, and consists of five main elements: input, weights, transfer function, activation function and output (Arslan et al., 2019).

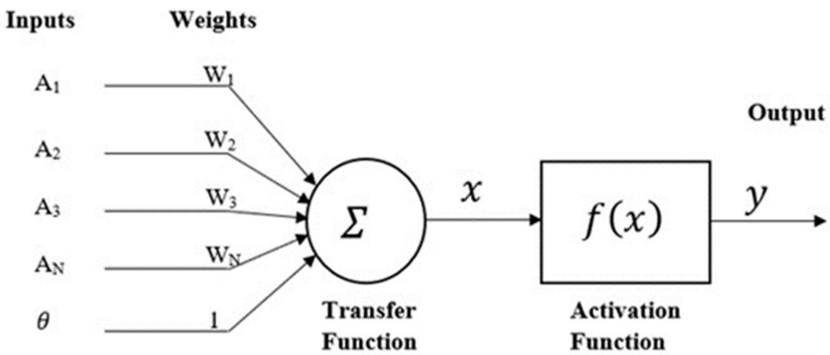

Figure 1. General Structure of a Neuron

In ANN models, a single hidden layer of feed forward back propagation structure is used. While the data flow in forward feeding is one-way without creating a loop between neurons, the error values in the back propagation are compared with real results and the weights at the input are updated (Gullu et al., 2011). The flow chart of the model formed for an ANN is given in Figure 2.

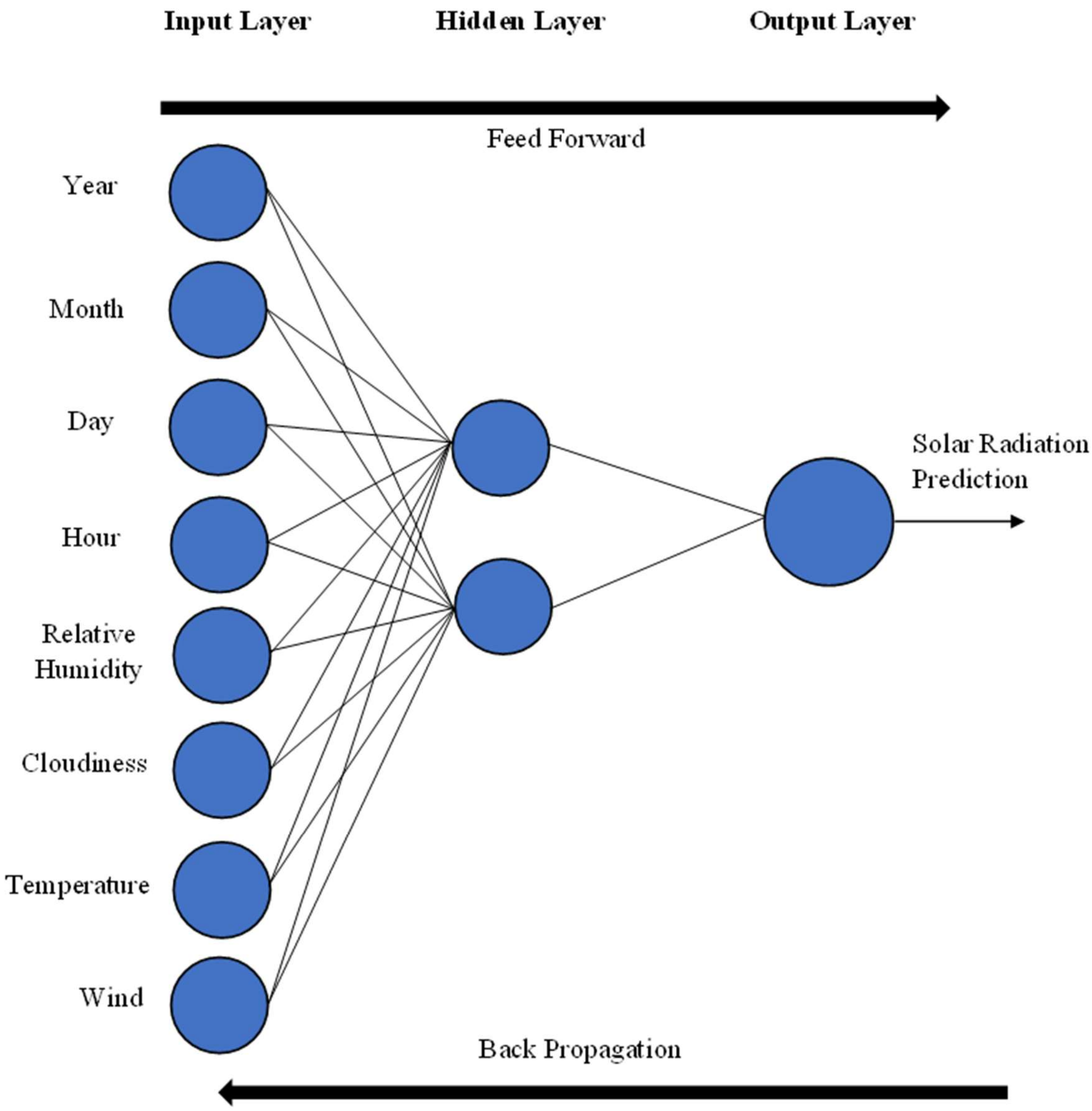

Figure 2. Created Feed Forward Back Propagation ANN Model

For ANN model in the study, Sigmoid function was used as an activation function.

2.2.4. Deep Learning (DL)

The DL algorithm is a subclass of ML based on ANN. In the DL, the entire process is carried out by the algorithm but in other ML algorithms the process needs to be modified manually in case of incorrect forecasting (Kayci, 2021). In this way, DL can better mimic the structure of learning and solving solutions to human learning algorithms compared to other ML algorithms. In addition, as shown in Figure 3, the increase in data in other ML methods supports the increase in performance up to a point, while the performance increase in DL continues with the amount of data (Döş and Uysal, 2019; Kraus et al., 2020).

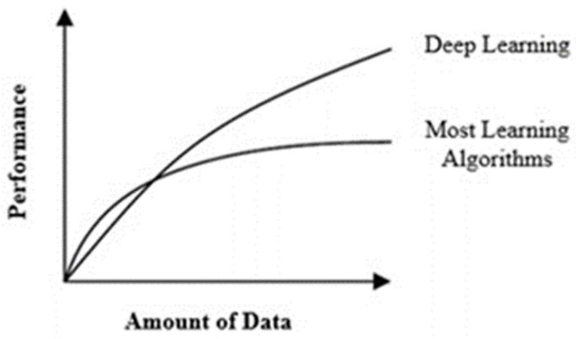

Figure 3. Performance Change According to the Amount of Data For DL And Other Learning Algorithms

In the DL model used, a structure consisting of a multi-layered feed forward backward propagation neural network was used with a stochastic gradient reduction. Also generated model has two hidden layers with 50 neurons in each hidden layer in addition to input and output layers. The generated model is given in Figure 4.

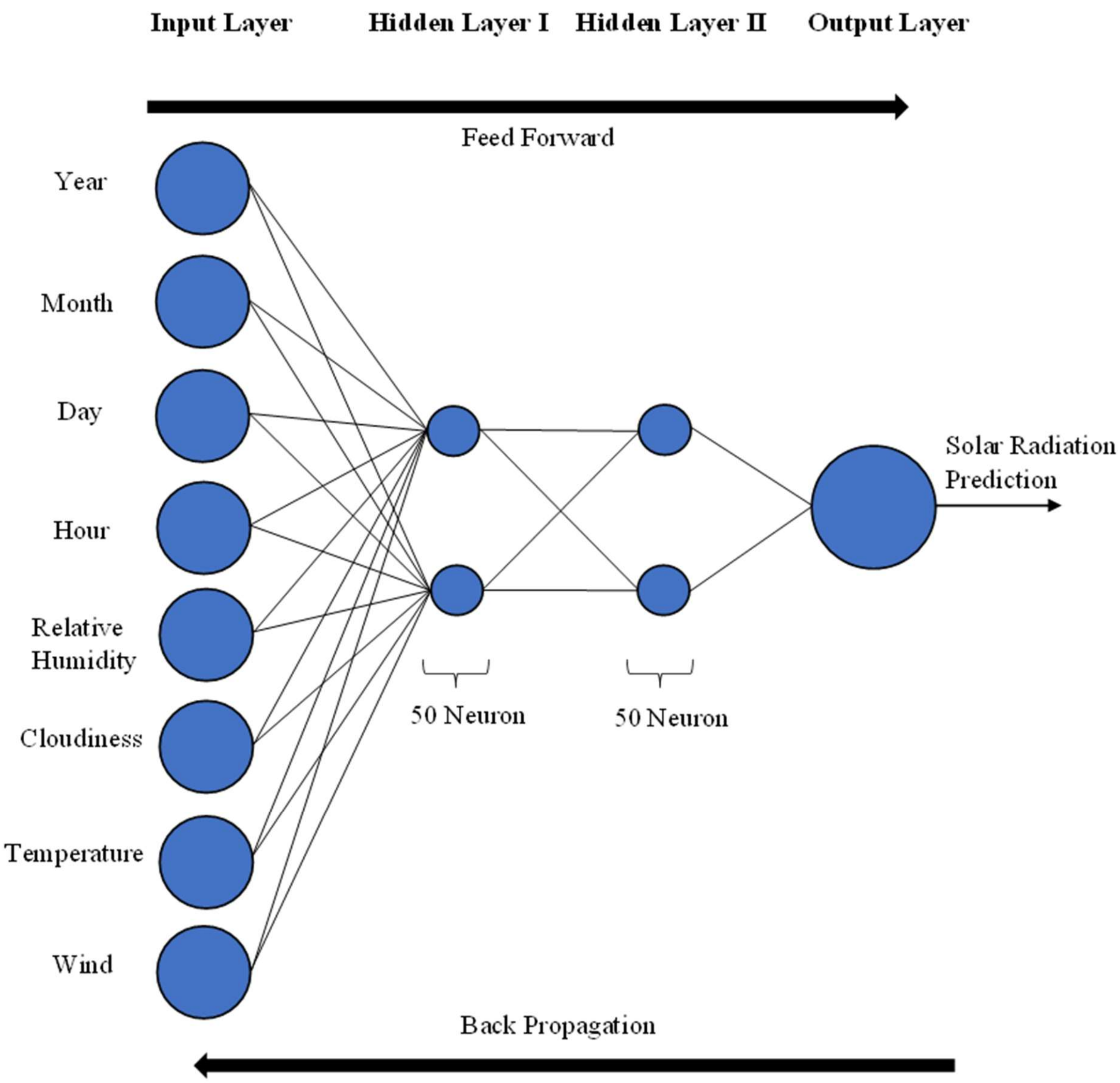

Figure 4. Designed DL model

For DL model in this study, Rectified Linear Unit (ReLU) was used as an activation function.

3. Results and Discussion

Eight different models were created for performance comparison by using dummy, numerical variables and four different algorithms. All models were tested with 5-fold cross validation. R-squared (R^2), The Root Mean Squared Error (RMSE) and Standard Deviation values obtained by cross validation are given in Table 3.

Table 3. Performance Results of the Models

Learning Algorithm	Time Variable Type	R^2	Standard Deviation	RMSE	Standard Deviation
RF	Numerical	0.929	0.004	4526.757	112.351
	Dummy Variable	0.801	0.002	9249.251	60.363
k-NN	Numerical	0.887	0.004	5655.029	80.441
	Dummy Variable	0.561	0.008	11145.828	86.704
ANN	Numerical	0.938	0.004	4266.399	115.200
	Dummy Variable	0.948	0.002	4024.826	153.142
DL	Numerical	0.942	0.003	4104.559	131.380
	Dummy Variable	0.947	0.003	3897.705	81.380

Accordingly, the most successful model in the RMSE value was the DL model with dummy variables. Right after, ANN model using dummy variables, comes next. The model with the worst performance values among all models was k-NN with dummy variables. Comparisons of the best and worst model's actual and forecasted values for a sample day (02.07.2018) are given in Figure 5. In addition, the effect of using dummy variables on the forecast of the same models is shown in Figure 5.

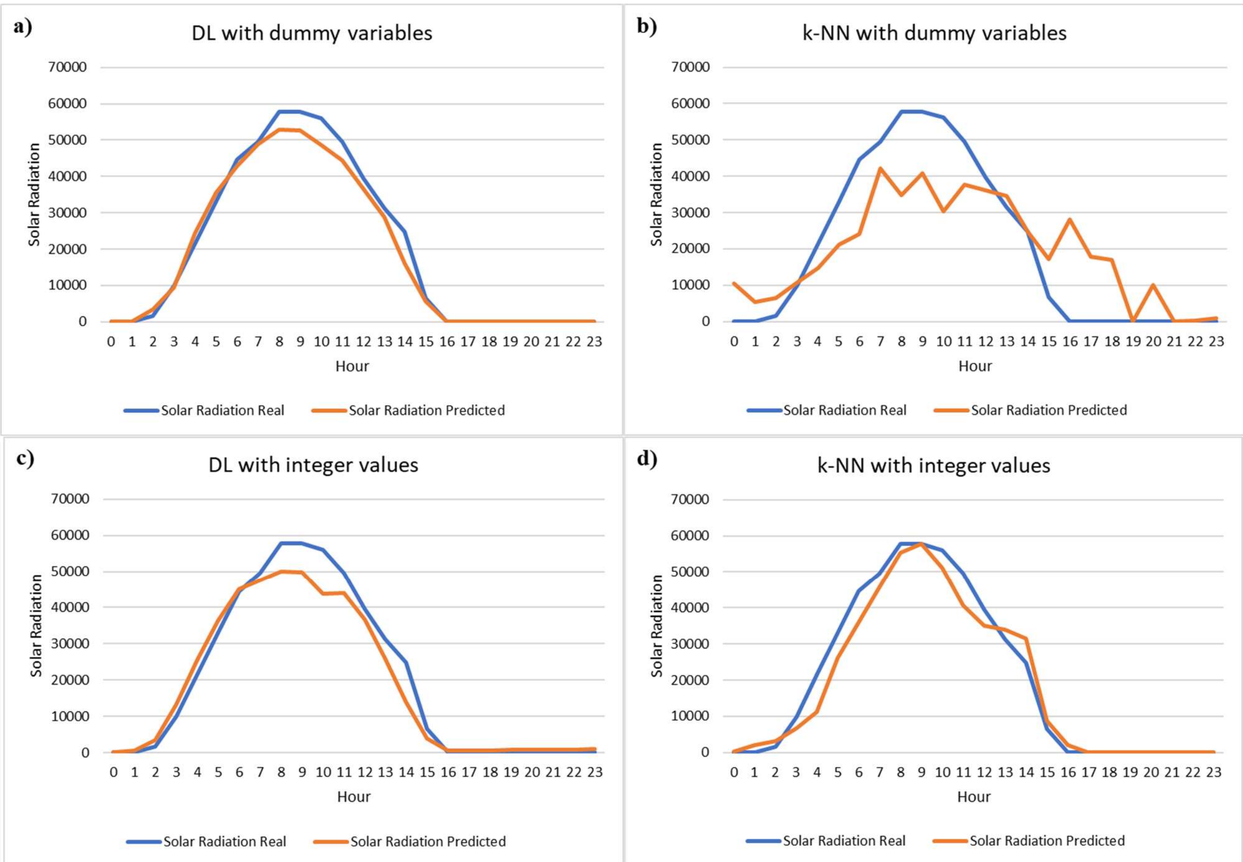

Figure 5. Comparison of the Actual and Forecasted Values For 02.07.2018. (a) DL with dummy variable, (b) k-NN with dummy variable, (c) DL with integer values, (d) k-NN with integer values.

The effect of using dummy variables on model performance is given in Table 4. Here, the RMSE variation describes the percentage change in the performance of the machine learning algorithm with the use of dummy variables. According to the performance results obtained, it was observed that the use of dummy variables decreased the success in the RF and k-NN models, while it increased the success in the ANN and DL models. The highest percentage performance increase occurred in the ANN model, and the highest performance loss occurred in the RF model. Enhancement on the performance of ANN and DL is notable. For ANN model 5.66% less RMSE value and 1% better accuracy has been obtained by using dummy coding.

Table 4. Effect of Dummy Variables on RMSE Values of the Models

Model	RMSE Variation (%)
ANN	-5.66
DL	-5.04
k-NN	97.10
RF	104.32

4. Conclusion

As the amount of electrical energy obtained from solar power plants increases, the forecasting of solar radiation gains importance day by day in order to ensure the balance of electricity production and consumption in the interconnected grid. For this reason, ML methods have become frequently used in this field. In the study, the data set collected from the meteorological data of Isparta province used and comparisons were made about the performance of ML algorithms in solar radiation forecast. In addition, the effect of using time variables as integer or dummy variables on the performance of the models was examined.

Among the models created in the study, the most successful model was the DL model using dummy variables with R^2=0.947 and RMSE=3897.705 (± 81.380) values. In addition, it was observed that the use of dummy variables decreased the performance in the RF and k-NN models and increased the performance in the ANN and DL models. The effect of using a dummy variable depends on the ML algorithm used.

In future studies, a more comprehensive model can be focused on by including data from different stations. In addition, different meteorological features can be added to the dataset, along with geographical parameters, and an increase in performance can be targeted with feature selection.

Acknowledgements

The authors thank the General Directorate of Meteorology for providing the data used in this study.

References

Alomari, M. H., Adeeb, J., & Younis, O. (2018). Solar Photovoltaic Power Forecasting in Jordan using Artificial Neural Networks. International Journal of Electrical and Computer Engineering (IJECE), 8(1), 497. https://doi.org/10.11591/ijece.v8i1.pp497-504

Al-Rousan, N., Al-Najjar, H., & Alomari, O. (2021). Assessment of predicting hourly global solar radiation in Jordan based on Rules, Trees, Meta, Lazy and Function prediction methods. Sustainable Energy Technologies and Assessments, 44, 100923. https://doi.org/10.1016/j.seta.2020.100923

Al-Sbou, Y. A., & Alawasa, K. M. (2017). Nonlinear Autoregressive Recurrent Neural Network Model For Solar Radiation Prediction. 12(14), 10.

Anonymus. (2021). "Renewable capacity highlights", International Renewable Energy Agency (IRENA).

Arslan, G., Bayhan, B., & Yaman, K. (2019). Mersin / Türkiye için Ölçülen Global Güneş Işınımının Yapay Sinir Ağları ile Tahmin Edilmesi ve Yaygın Işınım Modelleri ile Karşılaştırılması. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 7(1), 80–96. https://doi.org/10.29109/gujsc.419473

Ayko, O. & Bozkurt Keser, S. (2021). A comparison of machine learning algorithms for forecasting solar irradiance in Eskişehir, Turkey . International Journal of Applied Mathematics Electronics and Computers , 9 (4) , 103-109. https://doi.org/10.18100/ijamec.995506

Bamisile, O., Oluwasanmi, A., Ejiyi, C., Yimen, N., Obiora, S., & Huang, Q. (2022). Comparison of machine learning and deep learning algorithms for hourly global/diffuse solar radiation predictions. International Journal of Energy Research, 46(8), 10052–10073. https://doi.org/10.1002/er.6529

Benali, L., Notton, G., Fouilloy, A., Voyant, C., & Dizene, R. (2019). Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. Renewable Energy, 132, 871–884. https://doi.org/10.1016/j.renene.2018.08.044

Demirtas, M., Yesilbudak, M., Sagiroglu, S., & Colak, I. (2012). Prediction of solar radiation using meteorological data. 2012 International Conference on Renewable Energy Research and Applications (ICRERA), 1–4. https://doi.org/10.1109/ICRERA.2012.6477329

Demolli, H., Ecemiş, A., Dokuz, A. Ş., & Gökçek, M. (2019). Makine Öğrenmesi Algoritmalarıyla Güneş Enerjisi Tahmini: Niğde İli Örneği.

Döş, M. E., & Uysal, M. (2019). Uzaktan algılama verilerinin derin öğrenme algoritmaları ile sınıflandırılması. Türkiye Uzaktan Algılama Dergisi, 1(1), 28-34.

Elsheikh, A. H., Sharshir, S. W., Abd Elaziz, M., Kabeel, A. E., Guilan, W., & Haiou, Z. (2019). Modeling of solar energy systems using artificial neural network: A comprehensive review. Solar Energy, 180, 622–639. https://doi.org/10.1016/j.solener.2019.01.037

Erten, M. Y. & Aydilek, H. (2022). Solar Power Prediction using Regression Models . International Journal of Engineering Research and Development , Special Issue 2022 , 333-342. https://doi.org/10.29137/umagd.1100957

Gullu, M., Yilmaz, M., & Yilmaz, I. (2011). Application of back propagation artificial neural network for modelling local GPS/levelling geoid undulations: A comparative study. In FIG Working Week (pp. 18-22).

Hamdan, M. A., Abdelhafez, E., & Ghnaimat, O. (2017). Prediction of Hourly Solar Radiation in Amman-Jordan by Using Artificial Neural Networks. Int. J. of Thermal & Environmental Engineering, 14(2), 103-108.

Hong, T., Gui, M., Baran, M. E., & Willis, H. L. (2010, July). Modeling and forecasting hourly electric load by multiple linear regression with interactions. In Ieee pes general meeting (pp. 1-8). IEEE.

Ibrahim, I. A., & Khatib, T. (2017). A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. Energy Conversion and Management, 138, 413–425.

Jolly, S., & Gupta, N. (2021). Understanding and implementing machine learning models with dummy variables with low variance. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020, Volume 1 (pp. 477-487). Springer Singapore.

Kara, A. (2019). Uzun-Kısa Süreli Bellek Ağı Kullanarak Global Güneş Işınımı Zaman Serileri Tahmini . Gazi University Journal of Science Part C: Design and Technology , 7 (4) , 882-892 . https://doi.org/10.29109/gujsc.571831

KAYCI, B. (2021). Güneş Panellerinin Dört Rotorlu İha Kullanilarak Termografi Yöntemiyle Derin Öğrenme Tabanli Hata Tespit ve Teşhisi (Doctoral dissertation).

Korkmaz, D., Çelik, H. E., & Kapar, M. (2018). Sınıflandırma ve Regresyon Ağaçları ile Rastgele Orman Algoritması Kullanarak Botnet Tespiti: Van Yüzüncü Yıl Üniversitesi Örneği.

Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. European Journal of Operational Research, 281(3), 628–641. https://doi.org/10.1016/j.ejor.2019.09.018

Marzouq, M., El Fadili, H., Lakhliai, Z., Mechaqrane, A., & Zenkouar, K. (2019). New distance weighted k Nearest Neighbor model for hourly global solar irradiation estimation. 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 1–5. https://doi.org/10.1109/WITS.2019.8723697

Mohammed, L. B., Hamdan, M. A., Abdelhafez, E. A., & Shaheen, W. (2013). Hourly Solar Radiation Prediction Based on Nonlinear Autoregressive Exogenous (Narx) Neural Network. 7(1), 8.

Ozoegwu, C. G. (2019). Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number. Journal of Cleaner Production, 216, 1–13. https://doi.org/10.1016/j.jclepro.2019.01.096

Pardo, A., Meneu, V., & Valor, E. (2002). Temperature and seasonality influences on Spanish electricity load. Energy Economics, 24(1), 55-70.

Sanders, S., Barrick, C., Maier, F., & Rasheed, K. (2017). Solar Radiation Prediction Improvement Using Weather Forecasts. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 499–504. https://doi.org/10.1109/ICMLA.2017.0-112

Sun, H., Gui, D., Yan, B., Liu, Y., Liao, W., Zhu, Y., Lu, C., & Zhao, N. (2016). Assessing the potential of random forest method for estimating solar radiation using air pollution index. Energy Conversion and Management, 119, 121–129. https://doi.org/10.1016/j.enconman.2016.04.051

Şahin, M. (2013). Comparison of modelling ANN and ELM to estimate solar radiation over Turkey using NOAA satellite data. International Journal of Remote Sensing, 34(21), 7508–7533. https://doi.org/10.1080/01431161.2013.822597

Şeker, M. (2021). Yapay Sinir Ağı (YSA) Kullanılarak Meteorolojik Verilere Dayalı Solar Radyasyon Tahmini Estimation of Solar Radiation Based on Meteorological Data Using Artificial Neural Network (ANN). 13.

Torres-Barrán, A., Alonso, Á., & Dorronsoro, J. R. (2019). Regression tree ensembles for wind energy and solar radiation prediction. Neurocomputing, 326–327, 151–160. https://doi.org/10.1016/j.neucom.2017.05.104

Uğuz, S. , Oral, O. & Çağlayan, N. (2019). PV Güç Santrallerinden Elde Edilecek Enerjinin Makine Öğrenmesi Metotları Kullanılarak Tahmin Edilmesi . International Journal of Engineering Research and Development , Aralık 2019-Özel Sayı , 769-779 . https://doi.org/10.29137/umagd.514933

Yadav, A. K., Malik, H., & Chandel, S. S. (2015). Application of rapid miner in ANN based prediction of solar radiation for assessment of solar energy resource potential of 76 sites in Northwestern India. Renewable and Sustainable Energy Reviews, 52, 1093–1106. https://doi.org/10.1016/j.rser.2015.07.156

Yesilbudak, M., & Ozcan, A. (2022). k-NN Classifier Applications in Wind and Solar Energy Systems. 2022 11th International Conference on Renewable Energy Research and Application (ICRERA), 480–484. https://doi.org/10.1109/ICRERA55966.2022.9922701

Zeng, Z., Wang, Z., Gui, K., Yan, X., Gao, M., Luo, M., Geng, H., Liao, T., Li, X., An, J., Liu, H., He, C., Ning, G., & Yang, Y. (2020). Daily Global Solar Radiation in China Estimated From High-Density Meteorological Observations: A Random Forest Model Framework. Earth and Space Science, 7(2), e2019EA001058. https://doi.org/10.1029/2019EA001058

Zhou, Y., Liu, Y., Wang, D., Liu, X., & Wang, Y. (2021). A review on global solar radiation prediction with machine learning models in a comprehensive perspective. Energy Conversion and Management, 235, 113960. https://doi.org/10.1016/j.enconman.2021.113960