Research *Toylan/Kırklareli University Journal of Engineering and Science 8-1 (2022) 15-24*

DOI: 10.34186/klujes.1106357 *Geliş Tarihi:20.04.2022* *Kabul Tarihi:30.06.2022*

SOLAR IRRADIANCE PREDICTION USING BAGGING DECISION TREE-BASED MACHINE LEARNING

Hayrettin TOYLAN*[1]

[1] *Mechatronics Engineering Department, Faculty of Technology, University of Kırklareli, Türkiye*

Abstract

Solar energy is one of the most widely used renewable energy sources to generate electricity. However, the amount of solar radiation reaching the earth's surface is variable, creating uncertainty in the output of electrical power generation systems that use this source. Therefore, solar irradiance prediction becomes a critical process in planning. This study presents a short-term prediction of solar irradiance using bagging decision tree-based machine learning. As the inputs of the proposed method, air temperature, hour, day, month, and previous solar irradiance values were determined. The performance of the proposed method is tested on the measured data. The R^2 and RMSE values are 0.87 and 91.282, respectively, according to the results obtained. As a result, it has been revealed that the varying solar irradiance can be predicted with acceptable differences with this method.

Keywords: Renewable energy, Machine learning, Bagging decision tree, Solar irradiance prediction

TORBALAMA KARAR AĞACI TABANLI MAKINE ÖĞRENIMI KULLANARAK GÜNEŞ IŞINIMI TAHMİNİ

Öz

Yenilenebilir enerji kaynaklarından biri olan güneş ışınımlarının dünya yüzeyine düşen miktarının değişken olması bu kaynağı kullanan özellikle elektrik güç üretim sistemlerinin çıktısında belirsizlik yaratır. Bu nedenle güneş ışınımı tahmini planlamada çok önemli bir süreç haline gelmektedir. Bu makale, torbalama karar ağacı tabanlı makine öğrenimini kullanarak güneş ışınımının kısa vadeli bir tahminini elde etmeyi amaçlamaktadır. Önerilen yöntemin girdileri olarak hava sıcaklığı, saat, gün, ay ve önceki güneş ışınım değeri belirlenmiştir. Yöntemin performansı ölçülen veriler üzerinde test edilmiştir. Elde edilen sonuçlara göre R^2 ve RMSE değeri sırasıyla 0.87 ve 91.282 olarak bulunmuştur. Sonuç olarak bu yöntem ile değişen güneş ışınımlarının kabul edilebilir farklılıklarla tahmin edilebilir olduğu ortaya konmuştur.

Anahtar Kelimeler: Yenilenebilir enerji, Makine öğrenmesi, Torbalama karar ağacı, Güneş ışınımı tahmini

Sorumlu Yazar:Hayrettin TOYLAN, hayrettintoylan@klu.edu.tr

DOI: 10.34186/klujes.1106357 Geliş Tarihi:20.04.2022 Kabul Tarihi:30.06.2022

INTRODUCTION

Solar energy is one of the most important renewable energy sources when considering energy production costs. With the developing technology, these costs may decrease even further. Therefore, solar energy is critical in meeting the increasing energy demand every year [1]. The fact that global solar radiation varies greatly according to location and seasons creates the need for solar radiation prediction for the planning of energy to be obtained from this source [2]. This short-term forecast plays an important role in determining the power to be demanded from backup power plants to ensure the voltage and frequency stability of the power grid [3, 4].

The complexity of solar radiation prediction as an important and nonlinear problem motivates many researchers to come up with solutions to this issue. The forecasted solar radiation can be divided into three according to the time. Short-term forecast is the forecast for from one minute to several hours; medium-term forecast is one week to one year and long-term forecast is a forecast for more than a year. Each forecast range used serves a different task [5].

Corea et al. [6] developed a method for predicting solar radiation at variable time intervals of 1 to 6 hours of solar radiation using an artificial neural network (ANN) algorithm. As an algorithm input, temperature, humidity, pressure, wind and other forecasts obtained from the neighboring station were used. The results of the study show that they were able to predict short-term global solar radiation with error rates of less than 20% nRMSE. Aljanad et al. [7] forecasted solar radiation at intervals of several minutes or several seconds with artificial neural networks optimized for parameters (number of hidden layers, number of neurons, and learning rate) by particle swarm method. As method inputs, relative humidity, air pressure, aperture index, and average temperature were used. In the 5 seconds, they found RMSE at 1.7078, RMSE at 0.7537, MSE at 0.0292, and MAPE at 31.4348%. Feng et al. [8] compared the success of machine learning models and empirical models for global solar radiation estimation using only air temperature as input. In general, the hybrid mind evolutionary algorithm and artificial neural network model they proposed provided the best predictions. In another study, a convolutional neural network (CNN) and a bi-directional short-term memory (BILSTM) based hybrid deep learning (DL) model were proposed for medium-term solar radiation estimation. The R^2 value of this model was found to be 0.90 [9]. Long-term solar radiation forecasting comes to the fore in grid planning and design. Aslam et al.

DOI: 10.34186/klujes.1106357 Geliş Tarihi:20.04.2022 Kabul Tarihi:30.06.2022

compared different deep learning approaches to predict hourly and daily solar radiation from a year ago, and found that the closed unit GPU approaches with long short-term memory successful. In the proposed methods, past solar radiation data and open-sky spherical horizontal radiation (GHI) were used as inputs [10].

The main objective of this study is to investigate the capability of decision trees, which is one of the controlled machine learning methods of solar radiation prediction based on minimum meteorological factors. In this method, the deviation in simple tree structures and the variance in complex tree structures are the negative side of the method. In order to prevent these negative situations, two popular community learning methods are recommended: boosting and bagging [11]. In this study, decision tree regression was examined by the bagging method that uses air temperature, clock, day, month, and previous solar radiation values as inputs for forecasting from an hour in advance. The forecasted determination coefficient of the model (R^2) and mean square root (RMSE) were evaluated based on error. The rest of this article is drafted as follows: In Chapter 2 titled "Material and Method”, field of study, data and method are briefly presented. Chapter 3 titled "Results and discussion" presents the results of the bagged tree algorithm and their comparisons. Chapter 4 "Conclusion" is described in this section of the study.

2. Material and Method

2.1. Data

Hourly data obtained between 2018 and 2019 were used from the modular meteorological data station established at Kayalı campus of Kırklareli University in northwest Turkey. 8760x2 data, including air temperature and previous solar radiation values, were used during the training and testing phase. At the station's location, the average air temperature was 14.11 ^{0}C and the average solar radiation was found 157.88 W/m^2.

2.2. Decision tree regression

The decision tree is an important machine learning algorithm used in classification and estimation. In this study, decision tree regression was used because our dependent variable is constantly variable. The decision tree is based on a hierarchical decision scheme, which is likened to a tree with node structures. The tree consists of a root node that contains all the data, a series of internal

DOI: 10.34186/klujes.1106357 *Geliş Tarihi:20.04.2022* *Kabul Tarihi:30.06.2022*

nodes, and leaf nodes. Each node in the decision tree makes a binary decision that separates one class from another. This is usually done by moving the leaf down the tree until the node is reached [12].

The fact that decision trees are efficient, insensitive to missing values, and the decision tree is created once can be considered its advances [13]. However, with the deviation in tree structures, variance emerges as the negative side of the method. In order to eliminate this disadvantage, decision tree methods integrated with community methods are also preferred. The community method suggests a new tree structure created by combining multiple tree structures to suppress their weaknesses [11].

2.2.1 Bagging Trees

In Breiman [14] study, it presents that the data set selected in the training of a single decision tree will affect the success. Accordingly, if a similar data set is created with resampling and regression trees are grown without pruning and averaging, the variance component of the output error is reduced while also eliminating the problem of over-compatibility in a single tree [14, 15, 16]. The flowchart of the Bagging Trees is given in Figure 1.

DOI: 10.34186/klujes.1106357 *Geliş Tarihi:20.04.2022* *Kabul Tarihi:30.06.2022*

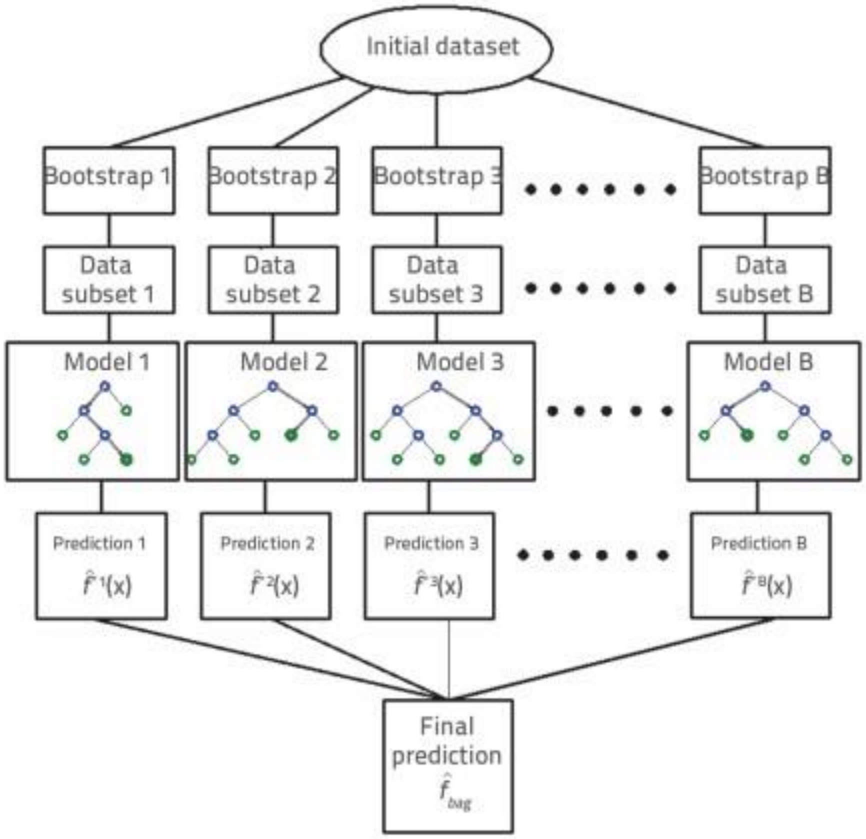

Figure 1. Flowchart of the Bagging Trees [17]

The first thing about bagging trees is to create a new training dataset the same size as the original data with the resampling method (Bootstrap) from the original training dataset. Then, each network created is trained separately in new training sets. In this study, 30 trees were used in bagging trees models. Finally, the average of all forecasts is calculated according to the Eq. (1) [18].

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} fi(X) \tag{1}$$

f_i represents each tree model. Bagging tree algorithm steps are given below [18]:

DOI: 10.34186/klujes.1106357 *Geliş Tarihi:20.04.2022* *Kabul Tarihi:30.06.2022*

Algorithm 1: Bagging trees approach

Input: Training and testing datasets, D
Output: Prediction outpout
for i=1,..., N *in TrainingDataset* **do**

- Take a bootstrapped replica D_i , from D
- Call Decision Tree with D_i and receive prediction $\widehat{y}_i$
- Add $\widehat{y}_i$ to the ensemble $\widehat{Y}$
- Compute the final prediction: $\widehat{\mathbf{y}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{y}_i$.

end
$Prediction_{BaggedTrees} \leftarrow \widehat{\mathbf{y}}$;
return $Prediction_{BaggedTrees}$

RESULTS and DISCUSSION

This section details the results of solar radiation prediction obtained by community tree-based regression with bagging method from machine learning methods. This study was modelled using Matlab software. Input parameters used in this regression method's training and testing phase are air temperature, pre-solar radiation value, time, day, and month information. The dependent variable of the method is determined as solar radiation. There are 8760 pieces collected hourly throughout the year from each data. No attribute selection algorithm has been applied to these attributes used in decision tree training. The features of the method used in the study are given in Table 1.

Table 1. Model properties

Model	Minimum number of leaf's	Number of learners	learning rate
Bagging Tree	8	30	1

Research *Toylan/Kırklareli University Journal of Engineering and Science 8-1 (2022) 15-24*

DOI: 10.34186/klujes.1106357 *Geliş Tarihi:20.04.2022* *Kabul Tarihi:30.06.2022*

The view of a part of our decision tree with five input features, which we obtained by the bagging method, is given in Figure 2.

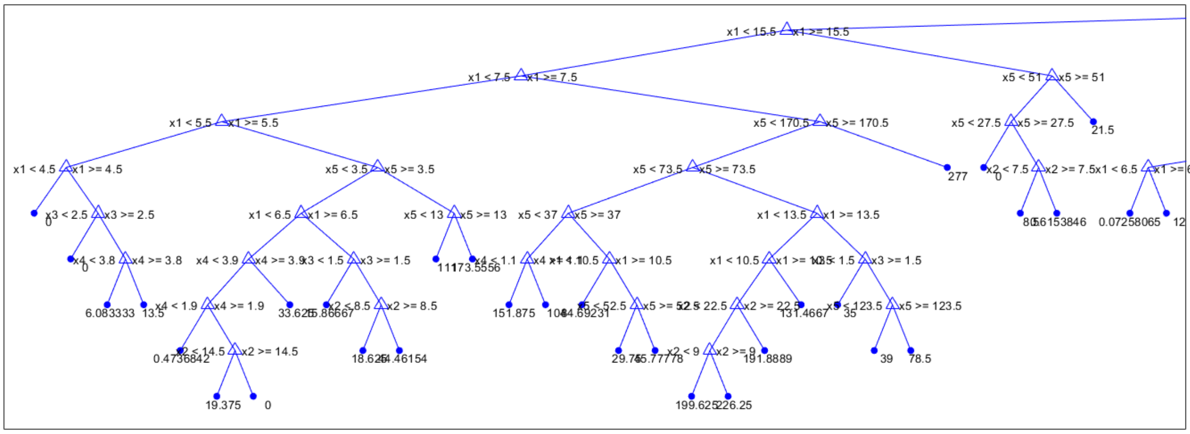

Figure 2. Part of the Bagging Decision Tree

The correct classification percentage (R^2) and the square root of the mean squared error (RMSE) are calculated to evaluate the forecast performance of the proposed method. These are some of the most commonly used methods to check the quality of the forecast. These are calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{t-1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t-1}^{n}(y_t - mean(Y))^2} \tag{3}$$

Here, *yt* is the measured values, *ŷt* are the forecasted values corresponding to the regression model, and n is the number of observations. The proposed model was compared with the traditional decision tree with a minimum leaf size of four (4). The R^2 and RMSE values obtained from the five (5) fold cross-validation are given in Table 2.

Table 2. Model success based on test data

Model	**R^2**	**RMSE**
Bagging tree	0.87	91.282
Decision tree	0.83	106.11

DOI: 10.34186/klujes.1106357 *Geliş Tarihi:20.04.2022* *Kabul Tarihi:30.06.2022*

The scatter plot is shown in Figure 3 to show the measured clearly and predicted solar radiation of the proposed model.

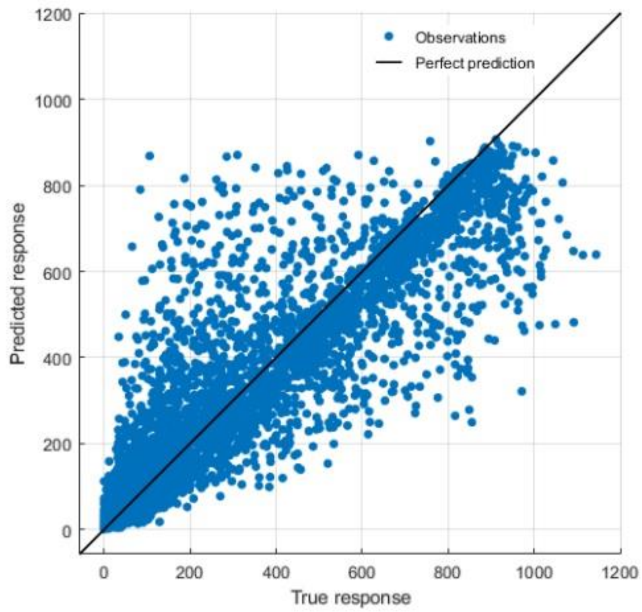

Figure 3. Scatter plot of the predicted response

In Figure 3, it can be seen that the bagging trees predict the solar radiation at an acceptably good level. Finally, in Figure 4, the 24-hour data on 16.03.2019 with the proposed method and compared with the measured values.

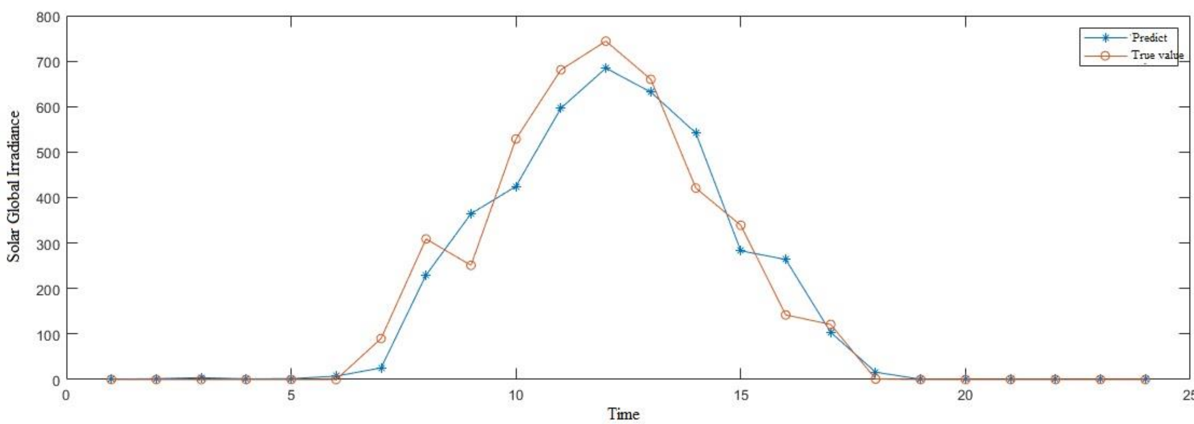

Figure 4. Comparison of true values and proposed method using the 24-hour datasets.

Thanks to the proposed method, it has been seen from the results that the solar radiation prediction is more successful than the traditional decision tree method. Furthermore, it has also been observed that the results obtained are at an acceptable level.

CONCLUSION

In this study, a bagging-based decision tree method is presented to realize the short-term prediction of solar radiation. Air temperature, previous solar radiation, time, day, and month information were selected as method inputs. R^2 and RMSE metrics were used to evaluate the performance of the proposed method. In addition, the proposed model is compared with a traditional decision tree prediction method. The results revealed that the use of bagged trees helps both reduce variance error and improve forecast quality. Future studies are planned to develop community learning methods using different input attributes to improve solar radiation forecasting performance.

REFERENCES

[1] Akarslan, E., & Hocaoglu, F. O. A novel method based on similarity for hourly solar irradiance forecasting. *Renewable Energy*, *112*, 337-346, 2017.

[2] Kamadinata, J. O., Ken, T. L., & Suwa, T. Sky image-based solar irradiance prediction methodologies using artificial neural networks. *Renewable Energy*, *134*, 837-845, 2019.

[3] Dong, N., Chang, J. F., Wu, A. G., & Gao, Z. K. A novel convolutional neural network framework based solar irradiance prediction method. *International Journal of Electrical Power & Energy Systems*, *114*, 105411, 2020.

[4] Tasnin, W., & Saikia, L. C. Deregulated AGC of multi-area system incorporating dish-Stirling solar thermal and geothermal power plants using fractional order cascade controller. *International Journal of Electrical Power & Energy Systems*, *101*, 60-74, 2018.

[5] Almonacid, F., Pérez-Higueras, P. J., Fernández, E. F., & Hontoria, L. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. *Energy Conversion and Management*, *85*, 389-398, 2014

[6] Gutierrez-Corea, F. V., Manso-Callejo, M. A., Moreno-Regidor, M. P., & Manrique-Sancho, M. T. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. *Solar Energy*, *134*, 119-131, 2016.

DOI: 10.34186/klujes.1106357 *Geliş Tarihi:20.04.2022* *Kabul Tarihi:30.06.2022*

[7] Aljanad, A., Tan, N. M., Agelidis, V. G., & Shareef, H. Neural network approach for global solar irradiance prediction at extremely short-time-intervals using particle swarm optimization algorithm. *Energies*, *14*(4), 1213, 2021.

[8] Feng, Y., Gong, D., Zhang, Q., Jiang, S., Zhao, L., & Cui, N. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. *Energy Conversion and Management*, *198*, 111780, 2019.

[9] Rai, A., Shrivastava, A., & Jana, K. C. A CNN-BiLSTM based deep learning model for mid-term solar radiation prediction. *International Transactions on Electrical Energy Systems*, *31*(9), e12664, 2021.

[10] Aslam, M., Lee, J. M., Kim, H. S., Lee, S. J., & Hong, S. Deep learning models for long-term solar radiation forecasting considering microgrid installation: A comparative study. *Energies*, *13*(1), 147, 2019.

[11] Lee, J., Wang, W., Harrou, F., & Sun, Y. Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. *Energy Conversion and Management*, *208*, 112582, 2020.

[12] Xu, M., Watanachaturaporn, P., Varshney, P. K., & Arora, M. K. Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, *97*(3), 322-336, 2005.

[13] Lu, H., & Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, *249*, 126169, 2020.

[14] Breiman, L. Bagging predictors. *Machine learning*, *24*(2), 123-140, 1996.

[15] Buhlmann P, Yu B. Analyzing bagging. Ann Stat 30:927–61, 2002.

[16] Prasad, A. M., Iverson, L. R., & Liaw, A. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems, 9(2), 181-199, 2006.

[17] Kovačević, M., Ivanišević, N., Petronijević, P., & Despotović, V. Construction cost estimation of reinforced and prestressed concrete bridges using machine learning. *Građevinar*, *73*(01.), 1-13, 2021.

[18] Harrou, F., Saidi, A., & Sun, Y. Wind power prediction using bootstrap aggregating trees approach to enabling sustainable wind power integration in a smart grid. *Energy Conversion and Management*, *201*, 112077, 2019.