Journal of Universal Computer Science, vol. 30, no. 6 (2024), 847-877
submitted: 5/7/2023, accepted: 8/1/2024, appeared: 28/6/2024

A Comparative Study of Data Mining Methods for Solar Radiation and Temperature Forecasting Models

F. Didem Alay
(Department of Computer Engineering, Harran University, Sanliurfa, Turkey
https://orcid.org/0000-0001-5711-9991, fdidemogretmen@harran.edu.tr)

Nagehan İlhan
(Department of Computer Engineering, Harran University, Sanliurfa, Turkey
https://orcid.org/0000-0002-1367-9230, nagehanilhan@harran.edu.tr)

M. Tahir Güllüoğlu
(Department of Electrical and Electronics Engineering, Harran University, Sanliurfa, Turkey
https://orcid.org/0000-0001-5165-6550, mtahir@harran.edu.tr)

Abstract: Photovoltaic (PV) energy systems are a leading type of renewable energy systems globally. Predicting PV energy production accurately is crucial for maintaining efficient energy grids, making informed decisions in the energy market, and reducing maintenance costs. To ensure high accuracy and optimal production, it is essential to monitor and analyze these variables regularly. Solar radiation and temperature are two meteorological variables that directly affect the quantity of PV energy generated in PV facilities. The Performance Ratio (PR) is a critical parameter for assessing PV plant performance. A comprehensive model was constructed in this study to forecast solar radiation and temperature using multiple machine learning methods, including Instance-Based K-Nearest Neighbor Algorithm (IBK), Linear Regression, Random Forests, Random Tree, Multilayer Perceptron (MLP), and MLP Regression. Moreover, we used time series approaches, such as Simple Exponential Smoothing (SES), Error-Trend-Seasonality (ETS), Autoregressive Integrated Moving Average (ARIMA) and Holt Winter's Seasonal Method (HWES) models for PV systems prediction. Initially, we conducted daily forecasts as well as 1-step ahead forecasts at 5-minute intervals for both solar radiation and temperature. It is crucial to subject both variables to the same methodology in order to construct precise models for forecasting PV. Secondly, we compared the predicted values of solar radiation and temperature with the actual energy yield of the power plant to calculate energy production. Subsequently, a relative analysis of data mining models and time series models have been performed depending on the statistical error criteria like RMSE, MAPE, MABE, MAE, MSE, and direction accuracy (DAC).

Keywords: Data Mining, Forecasting, PV Systems, Solar Radiation, WEKA, R
Categories: I.2, I.m, E.0, J.2
DOI: 10.3897/jucs.109080

1 Introduction

Photovoltaic (PV) systems are the most widely used alternative energy systems that transforming solar power into electrical energy. The rapid consumption of fossil energy sources in order to satisfy the energy demand is today increasing the need for the renewable energy sources that are unlimited and harmless to nature, ensuring the widespread use of renewable energy systems.

Today, energy consumption has increased to a large extent with the increasing population and industrial development, however, energy demand has not necessarily been met due to limited resources. The demand to use alternative energy sources (solar, geothermal, wind, hydroelectric, bioenergy, etc.) instead of conventional energy is increasing as a result of growing energy consumption, soaring costs, limited fossil fuel sources and greenhouse gas emissions. Therefore, many scientists are in search of unlimited renewable resources accessible all over the world with no harm to our nature. Although non-renewable energy sources are thought to be cheaper in the short term, renewable energy sources are unlimited in the long term and there is no cost of raw material. Photovoltaic energy is one of the best-known clean renewable energy resources like wind, solar, geothermal biomass, and hydraulic energy applications.

The planning and operation of energy systems have lately been handled from the “smart grid” perspective. Electric networks, called smart grids, involve the incorporation of novelties and services to make more dependable, cost-effective, and convenient applications for the environment. Therefore, increasing the utilization of renewable energy sources is one of the key goals. Due to the uncertain nature of such energy sources, advanced techniques are required to design and control the best smart grid. Therefore, forecasting methods provide valuable information for the efficient use of existing energy sources and the design of a safe and economical power system [AKCAN et al., 2020, Bracale et al., 2013]

Photovoltaic energy production is an innovative method of producing energy utilizing the photovoltaic effect generated by the semiconductor interface to transform solar energy into electrical energy directly. While renewable energy currently accounts for the 30% of the whole energy consumption, by 2030, photovoltaic energy production is expected to supply in excess of 25% of the world’s total energy need [Goswami and Sadhu, 2021, Zheng et al., 2011]. However, the capacity of PV power supplies installed in the world is expected to increase to 1721 GW by 2030 and 4670 GW by 2050 [Li et al., 2020].

PV power output changes depending on various such as varying climatic conditions, cloud coverage, pollution levels, solar panel surface [Loganathan et al., 2021] and solar radiation modelling errors. Solar radiation not only contributes to the generation of electricity but also increases the temperature of the PV system [Choo and Wei, 2022]. Therefore, forecasting models are used to take various factors into account for a reliable and efficient estimation of PV systems [Das et al., 2018]. Solar radiation and temperature are crucial meteorological variables that directly affect PV power generation. Proper analysis of these factors forms the basis for more accurate predictions. Cloud cover measurements are also one of the weather metrics that affect solar energy forecasts. However, measurements and forecasts of cloud coverage are frequently inaccurate, which limits the precision of day-ahead solar forecasts [Bashir et al., 2021]. Also, there are studies using numerical weather prediction (NWP) models to forecast solar radiation deterministically [Bashir et al., 2021].

Estimations observed from solar radiation and temperature prediction models are used in many fields such as agriculture, aviation and energy market. Horizons set as very short term (1 minute to several minutes ahead), short term (1-hour or several hours ahead to 1 day or 1 week ahead), medium term (1 month to 1 year ahead), and long term (1 year to 10 years ahead) are used to create forecasting models and different techniques are implemented in these forecasting studies [Raza et al., 2016]. Efforts to generate the most accurate and precise prediction models are still open to innovation and remain popular. Based on solar energy and temperature predictions, the architectural, scale, services and economic evaluation required for the installation of PV plants are made and the optimum site for the power plant is selected accordingly [Natarajan et al., 2021]

Forecasting solar radiation for PV systems using machine learning is a topic of interest in the field of renewable energy [Gaboitaolelwe et al., 2023, Viscondi and Alves-Souza, 2021]. These algorithms have demonstrated promising outcomes in precisely predicting solar radiation, which is important to optimize PV system performance and integrate PV power into the electrical grid. Accurate prediction of solar radiation and photovoltaic power generation is essential for controlling and managing electrical systems. [Cotfas, D., Marzband, M., Cotfas, P., Siroux, M., & Sera, 2022]. Forecasting solar radiation helps ensure that power demand can be met in stand-alone systems, smart grids, and grids. In this paper, we present a comprehensive model that incorporate both solar radiation and temperature variables, thereby creating and optimizing the estimation model for PV energy production systems. The proposed model estimates daily and 1-step ahead (for 5-minute intervals) solar radiation and temperature values. Different machine learning algorithms and time-series analysis methods have been used and compared to obtain accurate and meaningful information from the data used in modelling. Predicted solar radiation and temperature values are employed to calculate performance ratio (PR) and temperature-connected performance ratio (TPR). Our suggested model begins with data collection followed by the data input process, pre-processing data and analysis/forecasting phase, respectively. Solar radiation and ambient temperature data are estimated in the power prediction models for PV systems. Various accepted statistical error metrics such as MAPE, RMSE, MBE, MAE [Inman et al., 2013] were employed to assess the prediction accuracy. The efficiency of the forecasting methods was compared with the obtained forecast error. By using the predicted solar energy and temperature values, the daily and 1-step ahead energy production of the solar power plant were estimated, and the performance metrics of the predicted energy production were calculated by comparing it with the actual energy production values of the PV plant.

This paper is composed of five sections. Section 2 presents an extensive literature review to reveal the application of time series forecasting methods for renewable energy. Section 3 reviews the data mining concepts and presents the data collection, preprocessing, and analysis phases. Section 4 introduces the proposed solar radiation and temperature forecasting models and results. Finally, Section 5 is given with the conclusion of the paper and suggestions for future work.

2 Literature Review

Solar radiation and temperature values constitute the basis for designing the PV system. As far as the studies in the literature are concerned, forecasting models for solar radiation, temperature and PV power generation are to be developed for PV systems. PV power and energy produced by a solar system can be calculated through inputs such as estimated solar radiation values, temperature values, extent of the area, and effectiveness of solar panels [Mellit and Pavan, 2010].

Solar radiation forecasting models focus on solar and PV predictions for time horizons extending from a few minutes to a few days ago [Pelland et al., 2013, Sengupta et al., 2015]. The data acquired from meteorological reports, the PV system, satellites, sky observation [Andrade and Bessa, 2017, Diagne et al., 2013, Mathiesen et al., 2013, Mathiesen and Kleissl, 2011] as well as Numerical Weather Prediction-NWP models constitutes the basis of contemporary weather forecasts that are used in solar radiation and PV forecasts.

Solar radiation measurements are prone to mistakes compared to other meteorological variables and encounter technical or operational problems. For example, there may be problems with such issues as calibration, dirt sensors, accumulation of water and poles shading sensors. Solar tracking systems can also be used in order for the PV panels to absorb the maximum amount of solar radiation [Ben Othman et al., 2018]. Even the stations measuring global solar radiation may fail in recording the solar radiation data. Therefore, methods producing correct results must be developed so that a more accurate model and prediction for solar radiation can be achieved. There may be significant changes related with forecasting and balancing PV energy production at various timescales such as seconds, minutes and hours [Liu et al., 2019]. Also, it is important for a PV system to have a reliable and economical dedicated monitoring system in terms of obtaining energy yield, diagnosing system losses and performing performance assessments [Abdullah et al., 2020, Çiftçi et al., 2020].

Another factor in forecasting models for PV systems is temperature estimation. Temperature values are obtained from meteorological data. Meteorological data is commonly referred to as climate data or synoptic data. Synoptic data serves as real-time information for the purposes of aviation security and forecasting models. Conversely, climate data typically represents the formal record of data that has undergone qualitative evaluations. The temperature of PV modules can vary significantly, ranging from -100°C to +100°C on the surface of Earth [Libra et al., 2021]. The surface temperature of PV modules is influenced by environmental parameters including wind speed, ambient temperature, solar radiation, accumulated dust and relative humidity. [Azmi et al., 2023]. Additionally, the temperature distribution of PV modules in tropical climates can affect the system's output power [Rahman et al., 2023].

Considering the studies conducted on PV systems, data processing techniques are used to generate a data mining-based prediction framework in PV forecasting models. The available solutions could be categorized in three separate methodological sections; physical, statistical, and hybrid [Leva et al., 2017]. Physical models rely on mathematical equations defining the ability of PV systems to transform sources obtained from meteorological agents into electrical energy [Ogliari et al., 2017]. These models utilize numerical weather forecasts or satellite imagery to build accurate predictions [Xiang et al., 2021].

Today, techniques utilized the most for modelling the stochastic feature of solar radiation are statistical methods [Leva et al., 2017]. Accurately predicting solar radiation is crucial for forecasting the performance of PV systems. Various machine learning techniques, including support vector machines and artificial neural networks, have been implemented for effective solar radiation and energy production forecasting. [Fara et al., 2021, Gaboitaolelwe et al., 2023]. Techniques such as Regression Analysis [Sobri et al., 2018, Zafarani et al., 2018] , Time Series Analysis (Autoregressive Moving Average, Autoregressive Integrated Moving Average, Autoregressive Moving

Average With Exogenous Input, etc.) [Diagne et al., 2013, Inman et al., 2013, Yang et al., 2015], Genetic Algorithm [Chu et al., 2015], Pattern Recognition (KNN) [Antonanzas et al., 2016], Random Forests (RF) [Ahmad et al., 2018, Antonanzas et al., 2016], Extra-Trees (ET) [Ahmad et al., 2018], Support Vector Machine (SVM) [Antonanzas et al., 2016, Gaboitaolelwe et al., 2023, Zendehboudi et al., 2018] and ANN [Azadeh et al., 2013, Ding et al., 2011, Ramsami and Oree, 2015] analyse the historical dataset [Das et al., 2018] to estimate PV power. These models consider factors such as meteorological data and historical solar irradiance to improve the accuracy of the predictions. In this study, various forecasting models have been developed using many statistical methods for the approach of comparing these models. The models perform well in predicting time intervals ranging from a few minutes to several hours. Consequently, our forecasting models were designed to predict at 5-minute intervals. Non-linear methods such as Takagi-Sugeno (TS) fuzzy models and wavelet-based techniques outperform linear models, as demonstrated in [Leva et al., 2017]. Regression methods are commonly employed to describe intricate nonlinear atmospheric events after a preliminary estimate of a few hours. Artificial Neural Networks (ANNs) have been widely benefitted to estimate solar energy by means of meteorological fluctuations with regard to their capability to model nonlinear, dynamic, noisy data, and complicated systems. Specific calculation methods depending on ANN are employed to estimate the resulting energy amount of several hours.

Hybrid models are also used as a combination of two or more statistical and physical methods given in foresight studies above. Ramsami et. al. [Ramsami and Oree, 2015] presents a new technique to estimate stochastic energy output of photovoltaic systems 24 hours before their daily weather forecast. Comparing the performance of the new technique created by hybridizing Regression and ANN techniques with traditional Linear Regression and ANN models has also been reported. Kim et. al. [Kim et al., 2019] first developed a basic model that uses weather forecasts for the prediction of PV energy generation and then created an auxiliary model using a series of variables called auxiliary variables that were analyzed by weather agencies but not encapsulated in weather reports. Algorithms used in models are Artificial Neural Network (ANN), Linear Regression, Classification and Regression Tree (CART), Random Forest Regression (RFR), Support Vector Regression (SVR), Adaptive Boosting (AdaBoost), and k-Nearest Neighbors (k-NN).

Statistical estimation methods are generally the definition of historical data over time and the examination of the factors that cause variability. In other words, they are the investigation of causal effects. Statistical methods based on continuity and stochastic time series are used in PV energy prediction models. To predict solar radiation for a photovoltaic (PV) system using time series, several approaches can be considered. One approach is to use statistical models such as ARIMA and seasonal ARIMA (SARIMA) models [Fara et al., 2021]. These models incorporate short-term solar radiation predictions derived NWP models [Fara et al., 2021].In [Ji and Chee, 2011], a two-phase approach is used to estimate hourly solar radiation series. During the estimation phase, Autoregressive and Moving Average (ARMA) and Time Delay Neural Network (TDNN) are used to estimate stationary residual series. Time series analysis which especially includes modelling with ARIMA is extensively utilized in solar irradiance prediction areas [Sobri et al., 2018]. Yang et.al. [Yang et al., 2012] applies the time series analysis to devise an ARIMA model for the estimation of hourly solar radiation and cloud transients. In another study [Yang et al., 2015], Exponential

Smoothing method is used to estimate cloud cover effects for forecasting global horizontal irradiance (GHI) through polynomial regressions combined with knowledge-based heuristic time series decomposition methods in order to enhance computational efficiency and forecasting accuracy. In [Alsharif et al., 2019], SARIMA model is devised so as to estimate the solar radiation on a daily and monthly basis based on the hourly solar radiation data acquired using the Korean Meteorological Administration in the course of 37 years.

In the vast majority of previous studies in the literature, the solar radiation estimates and the temperature estimates from the meteorological data have been utilized separately. In this study, we develop a methodology that includes both solar radiation and temperature forecasting. Estimating solar radiation and temperature values within the scope of the same study will enable users to estimate solar power production faster and more accurately. In addition, it is also apparent that there is a lack of comparison by using so many different methods in the creation of forecasting models in the literature. In this study, it is aimed to eliminate this deficiency by using a variety of methods comparatively. The performance of methodology is evaluated by various metrics related to forecasting error. Thus, the combined use of solar radiation and temperature predictions, which play a key role in PV systems power generation, yields more successful results in terms of both cost and reliability.

3 Data Mining

In this study, we develop a methodology using forecasting models and optimize using data mining methods for PV systems, implementing and comprising forecasting models on existing PV systems by using the data obtained from the solar power plant and meteorological station of GAPYENEV Center at Harran University in Şanlıurfa, Turkey. The proposed methodology is applied to the data collected from GAPYENEV Center, yielding solar radiation and temperature predictions in return. Various statistical error metrics such as RMSE (root mean square error), MBE (mean bias error), MAPE (mean absolute percentage error), MAE (mean absolute error), nMBE (normalized mean bias error), nRMSE (normalized root mean square error), and some statistical calculations such as DAC (direction accuracy) are utilized to assess the accuracy of the predictions. The predicted results indicate that the error is expected to be minimal. First of all, error rates are evaluated in the prediction values made for the available data. If the method is successful, then predicting future values follow. However, if it is insufficient, a new model should be defined and the operations should be repeated. Fig. 1 shows the details of solar and temperature forecasting model process.

The proposed model in the Fig.1 includes three main stage. First, the observed data to be used is collected from the solar power plant and weather station in Data Input stage. After the data collection process, the data is preprocessed and analyzed in Data Preprocessing stage where the dataset is created by performing data cleaning and feature selection processes. Then, the dataset, which is split into training and test datasets, is transferred to Forecasting stage. In this stage, solar radiation and temperature are estimated on training and test dataset by developing models that are used statistical and time series analyzing methods.

3.1 Data Collection

The dataset harnessed in this study is historical weather and solar energy data collected from the solar power plant (Harran GES) and meteorological station of GAPYENEV center and is maintained by Harran University. The installed power of the solar power plant named Harran GES is 5,314.32 kWp. There are 5 inverters (ABB PVS800-57-1000kW-C) and 16,104 PV panels (LG Electronics LG330N1C-A5) in the power plant. The data contains weather and solar power plant observations measured in the 2018-2020 period at 5-minute intervals. The dataset has 245692 data (~13,5 MB - file size) that contains solar radiation, temperature, production, consumption, temperature, module temperature, humidity, and wind speed data to create prediction models.

In the dataset, solar radiation expressed in (W/m^2) represents irradiance that is the solar power per unit area. Production represents solar power production of solar power plant, consumption is the power consumption of the building receiving the produced energy from the solar power plant. Other data columns represent the observed meteorological data. Firstly, the dataset named gapyenev_santral_data was created and data mining methods were applied to this data set. 70% of the data in the data set was nominated as the training dataset and 30% as test dataset. A sample section taken from the dataset is given in Table 1.

3.2 Data Pre-processing

Data quality is a key issue in data mining. Data collected from the source will likely be incomplete, have inconsistencies, noise, or errors, and should be pre-processed to increase reliability. Otherwise, unreliable and noisy input data will yield incorrect results. Data pre-processing is a semi-automated, time-consuming data mining phase. The massive growth in the scale of data and consequently the need to analyze a massive amount of data in pre-processing necessitates more effective tools for automatic data pre-processing [Oğuzlar, 2003].

Among numerous data pre-processing methods, Data cleaning is applied to remove the noise in the data and to correct the inconsistencies. Data integration, another technique, combines multiple data sources in a suitable database by avoiding inconsistencies and redundancies in the data. Data transformations such as normalization can be applied specifically when distance-based algorithms are used. In data reduction, it is aimed to remove irrelevant, weakly relevant attributes or to reduce the data size by sampling. The correct data pre-processing techniques should be applied to enhance the quality of the data, thereby helping to improve the accuracy of the results obtained and/or time spent on data mining [Jiawei Han, 2012].

In this study, it is necessary to perform data cleaning operations on the dataset during the data pre-processing phase that is shown in Fig.1. Operations such as completing missing data, correcting noise to diagnose outliers and removing inconsistencies in data of the dataset are performed at the data cleaning phase. Thus, the dataset is prepared for applying data mining methods.

Datetime	Solar Radiation (W/m^2)	Power Production (kW)	Power Consumption (kW)	Temperature (°C)	Module Temperature (°C)	Humidity (g/m^3)	Wind Speed(m/s)
1.08.2018 12:00	958.94	3193.83	3245.87	31.7	51.9	18.0	3.2
1.08.2018 12:05	957.55	3209.33	3262.27	31.7	51.7	18.0	3.1
1.08.2018 12:10	962.99	3225.0	3277.1	31.5	51.3	18.0	2.5
1.08.2018 12:15	970.4	3219.38	3269.98	31.6	51.7	18.2	2.8
1.08.2018 12:20	966.24	3174.07	3224.66	31.6	50.9	18.0	2.3
1.08.2018 12:25	968.73	3187.95	3239.87	31.6	51.6	18.4	2.6
1.08.2018 12:30	966.3	3038.34	3083.91	32.1	51.3	18.6	2.6
1.08.2018 12:35	955.99	2824.73	2876.89	32.5	51.3	18.8	2.8
1.08.2018 12:40	954.38	2901.85	2958.76	32.5	51.1	19.1	3.7
1.08.2018 12:45	954.59	3171.71	3224.23	32.1	51.4	19.2	3.4
1.08.2018 12:50	956.44	3228.15	3281.14	32.0	51.5	19.4	3.2

Table 1: A sample of dataset

3.3 Feature Selection

Selecting the best featuring subset with the original dataset is the best definition of feature selection. In accordance with the algorithm used in the process, the assessment of the features yields the best k features among the n features in the dataset, making up the whole process of feature selection. With the aim of reducing the number of features in the dataset, feature selection includes the process of selecting the most helpful and significant features for the problem. Reducing the number of features provides such advantages in the analysis process as follow [BUDAK, 2018]:

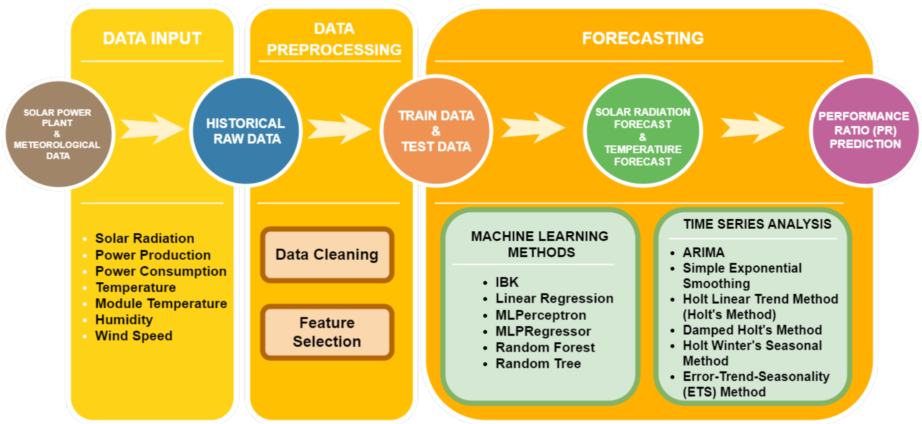

Figure 1: Solar radiation and temperature forecasting model process

- It saves resources in the data collection process required to create the data set,
- It reduces the required memory space for data storage,
- It increases the success of the proposed model,
- It diminishes the number of features and augments the pace of the algorithm,
- It eliminates irrelevant and noisy data,
- It improves data quality,
- It makes the data set easier to define, visualize and understand.

In this study, Correlation-based Feature Selection (CFS) which is a type of the filtering methods, is applied to the dataset at Feature Selection section in the model process, it is shown in Fig.1. CFS employs a searching algorithm in conjunction with a performance-evaluating function to assess the efficacy of feature subsets. The methodology employed by CFS for calculating the precision of feature subsets considers the predictive capacity of each feature with respect to the class label, as well as the internal correlation values among them. This methodology is founded on the premise that effective feature subsets are comprised of features with substantial correlation within their respective class [Hall, 1999]. CFS equation is given below:

$$M_s = \frac{k\bar{r}_{ci}}{\sqrt{k+k(k-1)\bar{r}_{ii}}} \quad (1)$$

In eq. (1), M_s indicates the correlation between the feature component and the outcome variable (the data) where k is the number of features in the subset, $\bar{r}_{ci}$ is the average correlation between M_s and the feature and $\bar{r}_{ii}$ is the average inter-correlation between the features.

4 Data Analysis and Forecasting Models

In this study, two popular software, WEKA [Hall et al., 2009] and R [Ihaka and Gentleman, 1996] are used to build models to make solar radiation and temperature predictions on dataset, comparing the accuracy to find the most efficient model among them. Fig. 1 shows the details of solar and temperature forecasting model process. In many different research fields such as agriculture, aviation, energy systems, solar radiation and temperature forecasting, models are developed at different forecasting horizons to ensure reliability and various methods are used in the estimation process.

In this study, the following time intervals employed to forecast radiation and temperature are as follow:

- 1-Day Ahead
- 1-Step Ahead (for forecasting horizon of 5-min.)

4.1 Model Performance Evaluation Criteria

To evaluate the effectiveness of forecasting models, it is necessary to employ multiple criteria for evaluation. The evaluation of performance regarding the accuracy and precision of models is performed with established statistical error indicators such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Bias Error (MABE), Mean Absolute Error (MAE), Mean Squared Error (MSE) and Direction Accuracy (DAC) metrics are listed in Table 3 to illustrate the disparities between the measured values and the estimated values. Besides, normalized metrics of standard statistical error indicators are also used in studies where prediction models are developed using time series analysis methods [Azimi et al., 2016, Kardakos et al., 2013, Zhang et al., 2014]. Therefore, normalized RMSE (nRMSE) and normalized MAE (nMAE) metrics are also applied in the accuracy assessment of radiation and temperature forecasting models.

Since the rate of variation between the seasonally observed data is high in the dataset normalized metrics are preferable. In the equations given in Table 2, y_i signifies the measured data, $\hat{y}_i$ signifies the predicted data and n signifies the instance number of the dataset.

4.1.1 Performance Ratio (PR)

For the developments in photovoltaic technologies to continue, it is very important that system performances are made correctly and consistently. Within the scope of IEC 61724 standards, the parameters to be used in assessing PV system performance are defined by the International Electrotechnical Commission [International Electrotechnical Commission, 1998]. Three of these parameters are used to determine the performance of grid-connected PV systems. These are the final yield, the reference yield and the performance ratio. [International Electrotechnical Commission, 1998]. Performance ratio (PR) is not an efficiency factor but a performance factor that is free from installed power and solar radiation intensity. Thus, it becomes possible to compare the PV power systems connected to the grid installed in various parts of the world with each other.

Model	Equation
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
Mean Absolute Percentage Error	$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|$
Mean Absolute Error	$MAE = \frac{1}{n}\sum_{i=1}^{n}\|y_i - \hat{y}_i\|$
Mean Squared Error	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
Normalized RMSE	$nRMSE = \frac{1}{y_{max}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \times 100\%$
Normalized MAE	$nMAE = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)}{y_{max}} \times 100\%$
Direction Accuracy	$DAC = \frac{1}{n}\sum_{i} 1_{sign(y_i - y_i - 1) = sign(\hat{y}_i - \hat{y}_i - 1)}$

Table 2: Evaluation metrics

Performance ratio PR formula is defined as (Equation (2)):

$$\mathrm{PR} = \frac{\mathrm{Y_f}}{\mathrm{Y_r}} = \frac{\mathrm{E}/\mathrm{P_{inst}}}{\mathrm{G}/\mathrm{G_0}} \tag{2}$$

Where $\mathrm{Y_f}$ is final yield that normalizes by taking the ratio of energy production (E) to the installed capacity ($\mathrm{P_{inst}}$) of the PV system. $\mathrm{Y_r}$ is the reference yield that normalizes the solar radiation (G) of a certain plane to the solar radiation ($\mathrm{G_0}$) under the Standard Test Condition (STC). In addition, to eradicate the impacts of temperature from the performance ratio, Temperature-corrected Performance Ratio (TPR) is used [Dierauf et al., 2013]. TPR formula is given below in Equation (3):

$$\mathrm{TPR} = \frac{\mathrm{PR}}{1+\gamma(\mathrm{T}-25)} \tag{3}$$

Where γ is the temperature coefficient of power and T is the module temperature. TPR is important for forecasting the output of a PV system since it minimizes the impact of weather changes over time points.

4.2 The Forecasting Models with WEKA

WEKA is an open source data mining program that was developed by Waikato University in New Zealand, incorporating machine learning techniques with a graphical user interface [Hall et al., 2009]. WEKA encapsulates various data pre-processing, regression, classification, association rules, clustering, and visualization. Algorithms can be applied directly to the dataset or by calling them from Java code. Furthermore, it is suited for devising novel machine learning algorithms.

In this study, using WEKA software, multiple methods are applied to develop solar radiation and temperature predictions with the dataset, and performances are compared. Data mining methods used in creating forecasting models with WEKA are:

Instance-Based K-Nearest Neighbor Algorithm (IBK): It is one of the simplest pattern recognition methods that classify objects based on the closest training examples in the attribute space. This algorithm classifies according to the class of the nearest neighbor as much as the given k-value. Classification of a vector in the IBK algorithm is done by using the known vectors [Kapur et al., 2017].

Linear Regression (LR): Regression is a statistical measurement that provides estimations according to the presence of a relationship between one dependent variable and other independent variables, tries to determine the strength of the relationship between these variables. Regression equations can help to determine whether the data fit an equation. The purpose of the regression model estimation is to find the regression line that best represents the relationship between the dependent variable and the predictive variables. Linear Regression is described as a line equation of the linear relationship among two variables, and provides predictions about the other if one of the values of the variable is given [Olive, 2017]. In order to correctly estimate the data, it is required to create the best line for the data. When creating the best line, the region closest to all points should be preferred. Since a line is created in Linear Regression, it represents only two studied variables (dependent and independent variable).

Random Forests (RF): As opposed to the conventional classification and regression trees, it creates a huge quantity of decision trees and grants the opportunity to assess the combination of these trees [Breiman, 2001]. In the RF method, the structure in which decision trees are built is referred to as a forest. In the forest, each decision tree is formulated through gathering samples from the data set using the bootstrap technique and determining the number of random variables gathered from all variables at each node.

Random Tree (RT): It is a classification algorithm that formulates a tree by taking a certain number of randomly selected features in each node [Aldous, 1991] without any pruning. It also has an option that allows estimation of class possibilities based on the acquired data set.

Multilayer Perceptron (MLP): This contains a classifier that uses back-propagation to sort WEKA samples. This network can be configured manually, created algorithmically or both. It can also be viewed and altered in the course of training time. The nodes in the network are sigmoid (except for one thing; when the class is numerical, the output nodes become linear units without thresholds) [Yang and Yang, 2014].

MLP Regressor (MLPR): It trains a multi-layered sensor with a single hidden layer using WEKA's optimization class. All features, including the target feature are standardized. It has several parameters such as the ridge parameter used to determine the penalty for the size of the weights. the sigmoid function is employed for classification in the output layer. If the approximate sigmoid is determined for hidden layers, it is also used for the output layer [Cornejo-Bueno et al., 2019].

4.3 Comparison of Results Obtained from Forecasting Models With WEKA

In this study, solar radiation and temperature forecast models are created by applying IBK, LR, MLPR, MLP, RF and RT methods separately in Weka and their results are presented in Table 3, Table 4, Table 5 and Table 6, respectively.

According to the performance criteria such as MSE, MAPE, MAE, RMSE and DAC it is evident that the best result of the daily solar radiation forecasting model is performed by the RT method. RT has the lowest value of error metrics overtaken by RMSE with 7,20, whereas MAE has the lowest with 0.18, MSE with 51.59 and MAPE with 256.86. DAC value is obtained as 78.55% from RT method. The same method has yielded the best results for the daily solar radiation predictions in direction accuracy, RMSE, MAE, MSE and MAPE with values 78.67%, 8.5, 3.36, 72.19 and 289.88 respectively.

Among the temperature forecasting models, RT method give the best results, similarly. For the 1-step ahead temperature forecasting model, RT method performs MAE value as 0.19, RMSE value as 0.26, MSE value as0,07 and MAPE value as 1.77 the lowest error values obtained from this method. DAC value is obtained as 43.69%. The best daily temperature forecasting results were produced by RT method where the direction accuracy, RMSE, MAE, MSE and MAPE values were obtained as 43.70%, 0.26, 0.07 and 1.70, respectively.

The results indicate that among the machine learning algorithms, RT method was provided the best performance. The RT method is a subset of the RF method. When the results are examined, our prediction models developed for solar radiation and temperature with the RF method also show very good performance. The RT method, by its nature, gives very good results in multidimensional and complex data sets. This result was determined once again when we applied the RT method to our solar power plant dataset containing complex data.

Method	**RMSE**	**MAE**	**nRMSE (%)**	**nMAE (%)**	**MAPE**	**MSE**	**DAC**
IBK	39.76	11.53	3.13	0.91	20.06	1581.12	85.73
Linear Regression	45.83	15.87	3.61	1.25	587.74	2100.30	44.73
MLPerceptron	46.21	18.81	3.64	1.48	1213.65	2135.43	44.72
MLPRegressor	44.79	14.42	3.53	1.13	578.76	2006.45	44.25
Random Forest	17.28	6.12	1.36	0.48	241.34	298.74	86.21
Random Tree	7.20	3.21	0.57	0.25	256.89	51.89	78.55

Table 3: 1-Step ahead - solar radiation forecasting models

Method	RMSE	MAE	nRMSE (%)	nMAE (%)	MAPE	MSE	DAC
IBK	38.47	10.95	3.03	0.86	17.11	1479.77	85.55
Linear Regression	46.02	16.43	3.62	1.29	514.68	2117.45	44.99
MLPerceptron	45.93	17.03	3.62	1.34	392.47	2109.52	44.50
MLPRegressor	45.69	15.74	3.60	1.24	133.80	2087.77	44.80
Random Forest	17.38	6.07	1.37	0.48	234.93	302.04	86.05
Random Tree	8.50	3.36	0.67	0.26	289.88	72.19	78.67

Table 4: Daily - solar radiation forecasting models

Method	RMSE	MAE	nRMSE (%)	nMAE (%)	MAPE	MSE	DAC
IBK	1.22	0.28	2.97	0.69	1.67	1.48	52.56
Linear Regression	2.03	0.54	4.94	1.30	3.01	4.11	43.41
MLPerceptron	1.47	0.50	3.60	1.22	3.45	2.18	43.64
MLPRegressor	17.86	0.46	43.53	1.12	3.18	3.19	42.41
Random Forest	0.58	0.22	1.41	0.53	1.57	0.33	45.75
Random Tree	0.26	0.19	0.64	0.46	1.77	0.07	43.69

Table 5: 1-Step ahead - temperature forecasting models

Method	RMSE	MAE	nRMSE (%)	nMAE (%)	MAPE	MSE	DAC
IBK	1.32	0.31	3.22	0.75	1.74	1.75	50.58
Linear Regression	2.05	0.51	5.01	1.25	3.05	4.22	43.55
MLPerceptron	1.61	0.59	3.93	1.44	5.11	2.59	42.63
MLPRegressor	2.04	0.55	4.97	1.34	4.16	5.15	43.42
Random Forest	0.62	0.22	1.52	0.55	1.56	0.39	45.10
Random Tree	0.26	0.18	0.63	0.44	1.70	0.07	43.70

Table 6: Daily - temperature forecasting models

Fig.2 shows the fit between the actual and the predicted output of 1-Step ahead solar radiation forecasting model with IBK method where the red colored plot area of the chart represents the predicted solar radiation values, the blue colored dots represent the actual solar radiation values by IBK method. Similarly, Fig.3 shows 1-Step ahead solar radiation forecasting model results with LR method while Fig.4 shows 1-Step ahead solar radiation forecasting model results with MLP method and Fig.5 shows the results with MLPR method. Fig.6 shows graphical representation of the fit between predicted and actual values obtained from forecasting model with RF method Fig.7 also shows this representation for RT method. As can be seen from the graph in Fig.7, the rate of similarity between the predicted and the actual values is high.

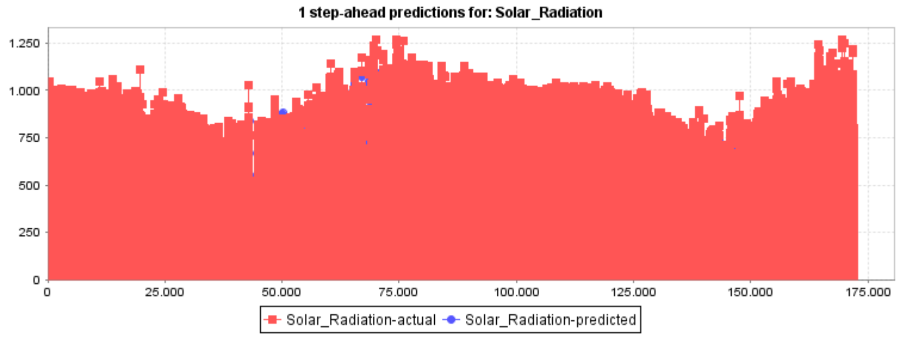

Figure 2: IBK solar radiation forecasting model

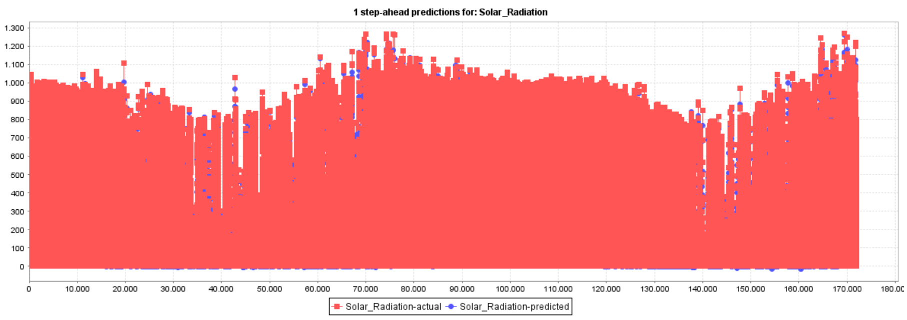

Figure 3: Linear Regression solar radiation forecasting model

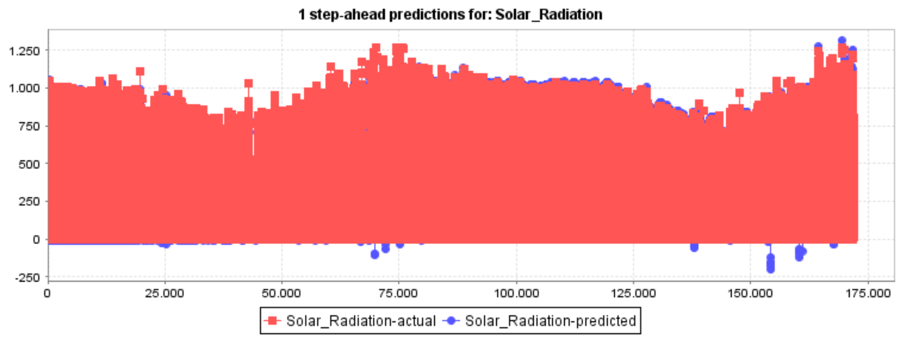

Figure 4: MLP solar radiation forecasting model

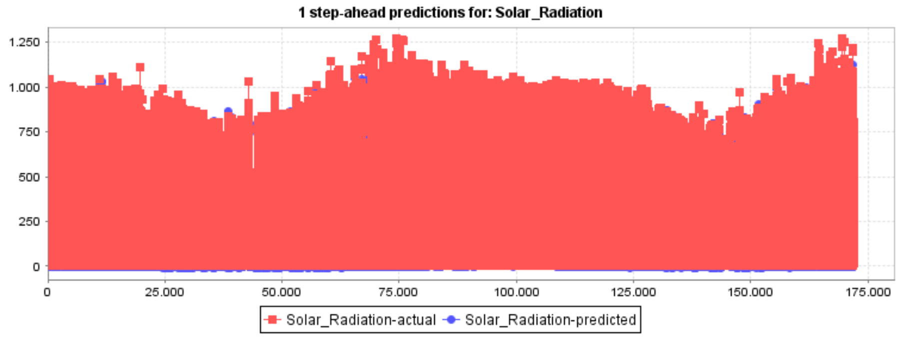

Figure 5: MLPRegression solar radiation forecasting model

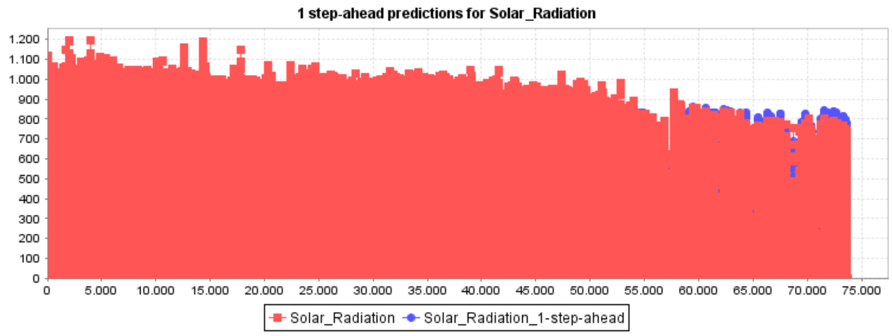

Figure 6: Random Forest solar radiation forecasting model

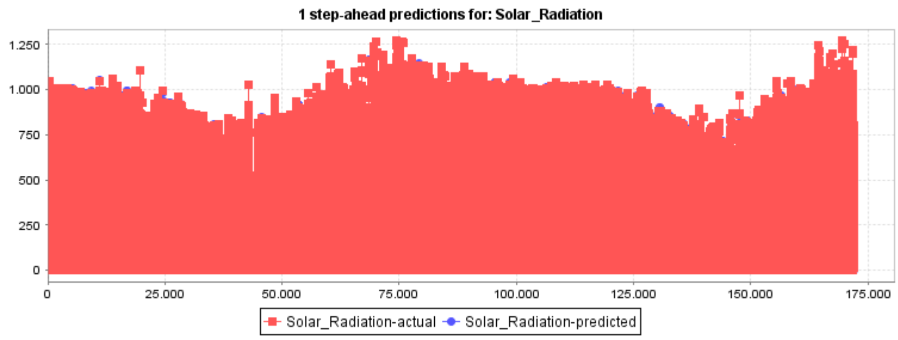

Figure 7: Random Tree solar radiation forecasting model

Similarly, Fig.8 shows the fit between the actual and the predicted output of 1-Step ahead temperature forecasting model with IBK method where the red colored plot area of the chart represents the predicted solar radiation values, the blue colored dots represents the actual solar radiation values by IBK method. Fig.9 shows 1-Step ahead temperature forecasting model results with LR method while Fig.10 shows 1-Step ahead temperature forecasting model results with MLP method and Fig.11 shows the results with MLPR method. Fig.12 shows graphical representation of the fit between predicted and actual values obtained from forecasting model with RF method Fig.13 also shows this representation for RT method. As can be seen from the graph in Fig.13, the rate of similarity between the predicted and the actual values is high.

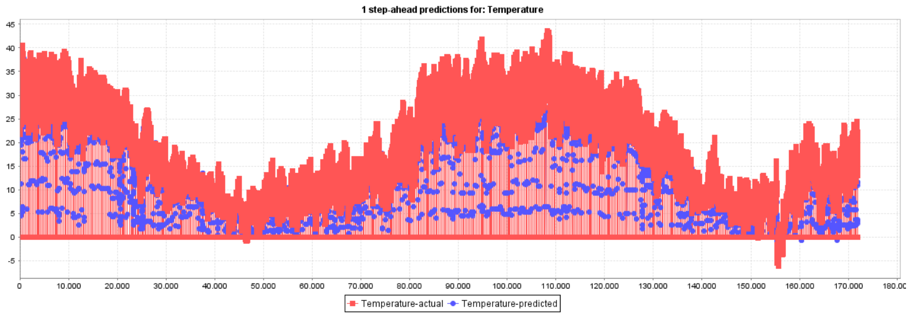

Figure 8: IBK temperature forecasting model

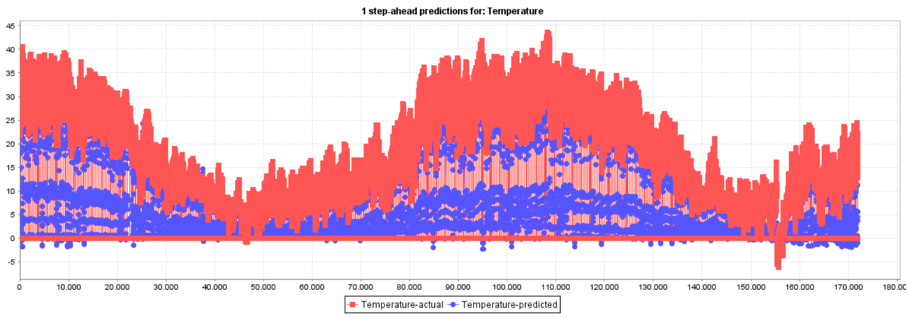

Figure 9: Linear Regression temperature forecasting model

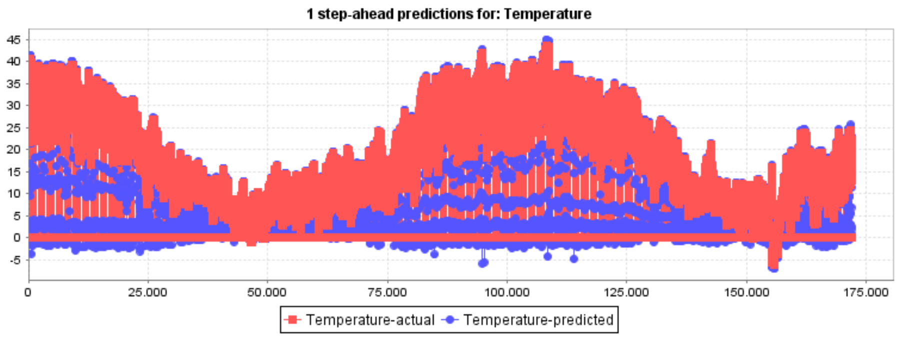

Figure 10: MLP temperature forecasting model

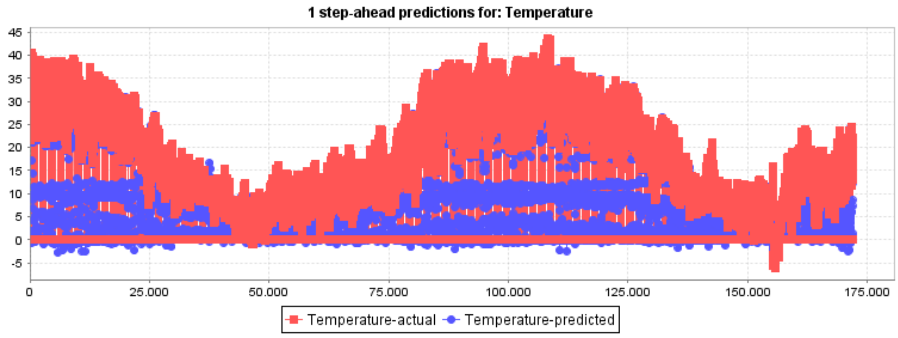

Figure 11: MLP Regression temperature forecasting model

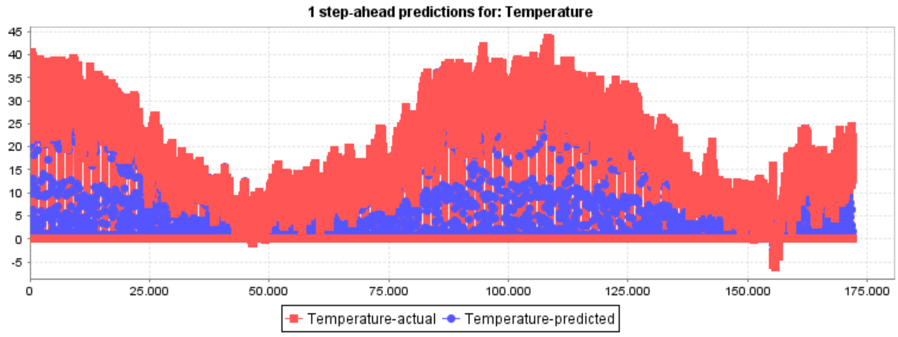

Figure 12: Random Forest temperature forecasting model

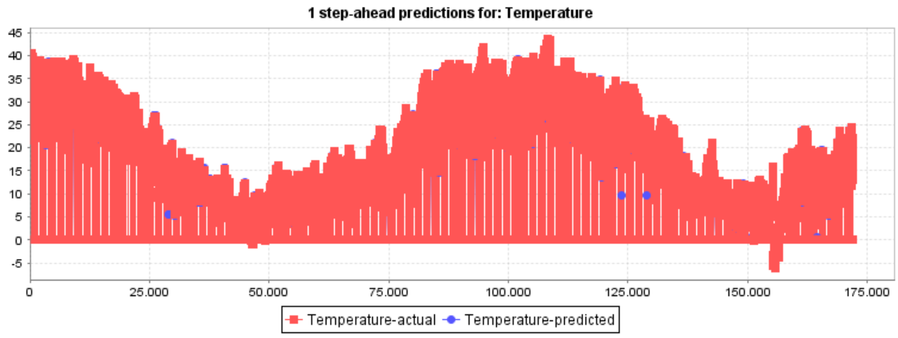

Figure 13: Random Tree temperature forecasting model

4.4 The Forecasting Models with R

In this study, we utilized time series analysis, which is a widely used statistical method in forecasting studies, for the development of solar radiation and temperature forecasting models. R programming language is used with R-Studio software for its application.

R is a customized, open source programming language for computing statistics and graphics. It was developed by John Chambers and his colleagues at Bell Labs (previously AT&T, now Lucent Technologies) and is modeled after the S language and environment. R is extensively used in the field of data analysis, offering tools for linear and nonlinear modeling, classical statistical tests, time series analysis, classification, clustering, and more [Ihaka and Gentleman, 1996].

R-Studio is a code-based program and basically runs through the R package program. In this study, R-Studio, a relatively more practical process, is used for modelling with time series analysis methods.

4.5 Time Series Analysis

A time series denotes a series of observations of a particular value arranged chronologically. The intention behind this analysis pertaining to the time series is to comprehend the level of accuracy conveyed by the set of observations and to accurately predict the future values of the variables within the time series. Time series forecasting is a conceptual framework utilized to predict future occurrences based on prior known events.

Time series comprises of four constituents;

1. *Trend component* refers to the stable state that ensues after the declining and ascending processes. Time series typically exhibit a long-term tendency towards decline or ascent.
2. *Seasonal component refers* to the fluctuations due to seasons in the time series. Certain periods within the series may differ from others due to seasonal factors.
3. *Cyclic Component* is a variation of a time period that is not related to seasonal variations.
4. *Irregular Component;* like other elements, they are unspecified variations that can be expressed by the term error.

After the time series are decomposed into all these constituent components, the Y time series can be expressed as the sum of the components $Yt = Tt + St + Ct + It$ or $Yt = TtStCtIt$ by the multiplication method.

Several methods exist for analyzing time series, the most widely known and used is the Box-Jenkins method for linear time series. It is successful in linear and stationary processes or time series that can be made stationary with several statistical methods. Nevertheless, since many time series have both linear and nonlinear relationships, alternative techniques are necessary for modeling the nonlinear relationships. ANN, which have the ability to model both linear and nonlinear relationships depending on the activation function within its structure, have emerged as a viable alternative method for analyzing time series data in recent years.

Box-Jenkins time series models available for estimation are Autoregressive (AR) model, Moving Average (MA) model, Autoregressive-Moving Average (ARMA) model, and Autoregressive Integrated Moving Average (ARIMA) model.

In this study, we performed time series analysis using ARIMA, SES, HOLT, HWES, and ETS methods to develop solar and temperature forecasting models.

ARIMA (Autoregressive Integrated Moving Average): These are the models applied to the series which are not stationary but converted into a stationary state by taking the difference. The models applied to the series that are stationary but converted into a stationary state by taking difference are called “non-stationary linear stochastic models”. These models are applied to the series with the difference of d degrees, where the value of the variable in the t period is expressed as a linear function of the error term in the same period with a certain number of back period values as well as the linear term of the error term in the same period and a certain number of the error terms in the t period. It is a combination of Moving Average (MA) models. The general representation of the models is $ARIMA\ (p, d, q)$. Here, p and q are the degree of Autoregressive (AR) Model and Moving Average (MA) Model where d is the degree of difference.

The general $ARIMA\ (p, d, q)$ model is formulated in the Equation (4):

$$z_t = \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + .. + \Phi_p z_{t-p} + \delta + \alpha_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2} - .. - \Theta_q a_{t-q} \tag{4}$$

Here $Zt, Zt-1, \ldots\ldots, Zt-p$ are the observation values taken from d degrees, $\Phi_1, \Phi_2, \ldots\ldots \Phi_p$ coefficients for the observation values taken from d degrees, δ is constant value, $\alpha_t, \alpha_{t-1}, \ldots, \alpha_{t-q}$ shows error terms and $\Theta_1, \Theta_{2,}.., \Theta_q$ shows coefficients for error term [Kırbaş, 2018].

Apart from Box-Jenkins methods, the methods frequently used in time series analysis are given below.

SES (Simple Exponential Smoothing): Exponential Smoothing method emerged in the late 1950s and pioneered some of the most successful forecasting methods. Estimates produced using exponential correction methods are the weighted averages of past observations and the weights drop excessively as observations grow. In other words, the closer the observation gets, the higher the associated weight becomes. These methods are most effective when the parameters that define the time series change slowly over time.

SES method is used to predict the time series when there is no clear trend or seasonal component, whereas the average (or level) of the y_t time series changes slowly over time.

$$\hat{y}_{t+1} = a y_t + (1 - a)\hat{y}_t \tag{5}$$

HOLT (Holt’s Linear Trend Method): Holt (1957) developed this method by expanding SES method to ensure data is predicted with a trend. The method includes a prediction equation and two correction equations (one for the level and the other for the trend):

$$Forecast \quad \hat{y}_{t+h}|_t = l_t + h_{bt} \tag{6}$$

$$Level \quad l_t = a y_t + (1 - a)(l_{t-1} + b_t - 1) \tag{7}$$

$$Trend \quad b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)_{b_{t-1}} \tag{8}$$

HWES (Holt-Winter's Seasonal Method): Holt (1957) and Winters (1960) created this method by extending Holt's seasonal capture method. Two types of Holt-Winters methods are designed for time series with linear trend:

Additive (Holistic) Holt-Winters method: Used for time series with constant seasonal variations.

$$\hat{y}_{t+h}|_t = l_t + h_{bt} + s_{t+h-m(k-1)} \tag{9}$$

$$l_t = a(y_t - s_{t-m}) + (1-a)(l_{t-1} + b_{t-1}) \tag{10}$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)_{b_{t-1}} \tag{11}$$

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \tag{12}$$

Multiplicative Holt-Winters method: Used for time series with multiplicative seasonal variations.

$$\hat{y}_{t+h}|_t = (l_t + h_{bt})s_{t+h-m(k-1)} \tag{13}$$

$$l_t = a\frac{y_t}{s_{t-m}} + (1-a)(l_{t-1} + b_{t-1}) \tag{14}$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)_{b_{t-1}} \tag{15}$$

$$s_t = \gamma\frac{y_t}{l_{t-1} - b_{t-1}} + (1-\gamma)s_{t-m} \tag{16}$$

Holt-Winters method is an exponential correction approach for seasonal data processing. Multiplicative Holt-Winters method is better known than others.

ETS (Error-Trend-Seasonality Method): This method, the foundations of which were laid by Pegels, was developed as a new approach by Makridakis, Andersen, Carbone, Fildes, Hibon, Lewandowski, Newton, Parzen and Winkler (1982). Then, the exponential smoothing method was arranged to obtain the optimum model with trend and seasonal components. The method was later expanded by Hyndman and Athanasopoulos (2009) by bringing 30 additive and multiplicative models together. Similar to the ARIMA method, the best model is selected among these models with the help of Akaike's Information Criterion (AIC) and the error series of the model assumes the white noise assumption. Considering the differences in the additive and multiplication methods within the trend and seasonal components, the combinations of the exponential smoothing methods for the ETS method are listed as follows:

Error: Additive (A), Multiplicative (M),

Trend: None (N), A, decreasing additive (Name), M, Decreasing multiplicative (Md),

Seasonal: N, A, M

In this study, time series analysis methods were applied by using RStudio to develop solar radiation and temperature forecasting models on the dataset. 70% of the dataset is separated as a training set and 30% is set as a test dataset, and prediction phases are performed with the time series analysis methods described above. The statistical performance criteria are calculated for each method and the accuracy of the models is compared.

To perform forecasting models with the time series analysis methods above, the dataset was first converted into the appropriate time series form. Forecasting models are developed for two different forecasting horizons;

- 1-Day Ahead
- 1-Step Ahead (for forecasting horizon of 5-min.)

4.6 Comparison of Results Obtained from Forecasting Models With R

In Table 7, the comparison of the error rates of all the time series methods (ARIMA, SES, HOLT, HWES and ETS) used to develop solar radiation forecasting models is given. The best solar radiation forecasting results for 1-day ahead were produced by Damped HOLT's method where the nRMSE and nMAE values are obtained as 13.06% and 5.12%. When the results are compared, although the given error rates of the solar radiation forecasting models by these time series analysis methods are very close to each other, the accuracy and the stability of the performance of the Damped Holt's model is higher more than others.

Method	**RMSE**	**MAE**	**nRMSE (%)**	**nMAE (%)**
ARIMA	42.10	14.94	5.16	1.83
Simple Exponential Smoothing	42.43	14.05	5.19	1.72
Holt Linear Trend Method (Holt's Method)	42.35	13.72	5.17	1.67
Damped Holt's Method	42.15	13.06	5.12	1.59
Holt Winter's Seasonal Method (Additive)	42.35	13.82	5.18	1.69
Error-Trend-Seasonality (ETS) Method	42.54	14.55	4.88	1.67

Table 7: Evaluation of solar radiation forecasting models with time series method

Fig.14 – Fig.18 show the comparison graphs between the actual and the predicted output of 1 day ahead solar radiation forecasting models with ARIMA, Damped HOLT's, SES, HWES and ETS method respectively where the black colored curve with dots represents the actual solar radiation values, the red colored area represents the fit of predicted solar radiation values by these methods.

When the graphs represented on the figures for all the time series methods given above, even though the curve of solar radiation time series harbors a strong seasonality pattern, it shows some strong cyclical behavior. There is no clear trend in the data during this period. However, the best fit among the actual and predicted values is obtained from the solar radiation forecasting model developed with Damped HOLT's method.

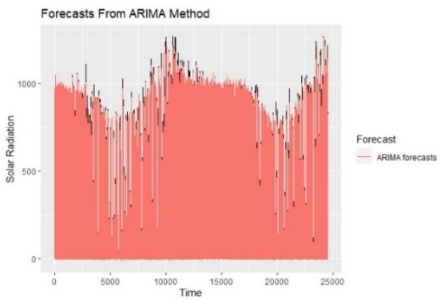

Figure 15: ARIMA solar radiation forecasting model

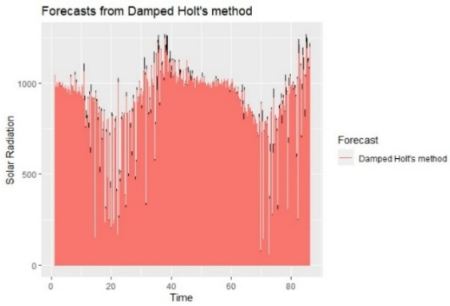

Figure 14: Damped Holt's Method solar radiation forecasting model

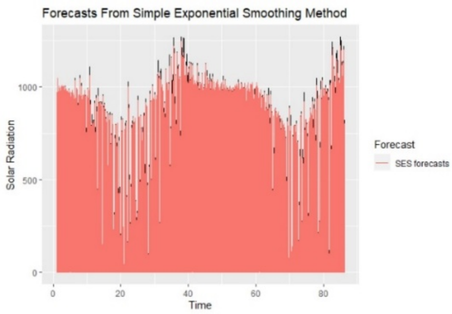

Figure 16: SES solar radiation forecasting model

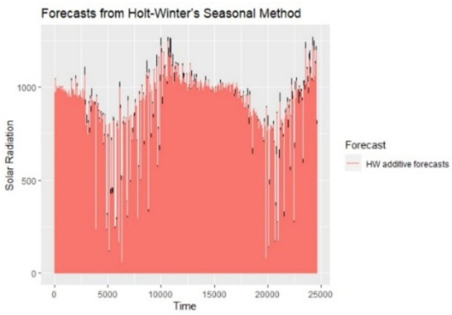

Figure 16: HWES solar radiation forecasting model

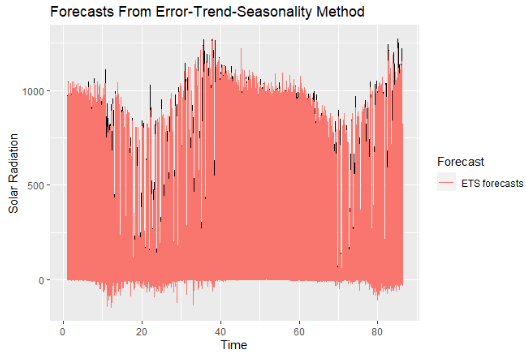

Figure 17: ETS solar radiation forecasting model

In the Table 8, the evaluation results of the time series models for temperature are given. It can be seen that all of the error rates are very close to each other, when evaluated in detail, temperature forecasting model with ARIMA method gives the lowest values of nRMSE and nMAE as 11.73% and 2.75%.

Method	**RMSE**	**MAE**	**nRMSE (%)**	**nMAE (%)**
ARIMA	2.61	0.61	11.73	2.75
Simple Exponential Smoothing	2.63	0.59	11.81	2.65
Holt Linear Trend Method (Holt's Method)	2.63	0.59	11.80	2.65
Damped Holt's Method	2.63	0.59	11.81	2.65
Holt Winter's Seasonal Method (Additive)	2.63	0.61	11.79	2.73
Error-Trend-Seasonality (ETS) Method	2.81	0.85	12.76	3.84

Table 8: Evaluation of temperature forecasting models with time series methods

Fig.19 – Fig.23 (a) shows 1 day ahead temperature forecasting model results with ARIMA, Damped HOLT’s, SES, HWES and ETS method respectively where the black colored curve with dots represents the actual solar radiation values, the red colored area represents the fit of predicted temperature values by these methods. The best fit among the actual and predicted values is obtained from the temperature forecasting model developed with ARIMA method.

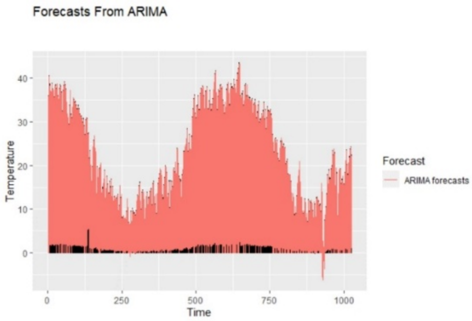

Figure 18: ARIMA temperature forecasting model

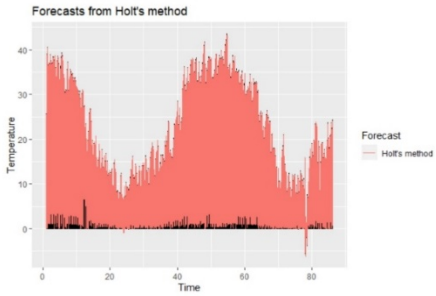

Figure 19: Damped Holt's Method temperature forecasting model

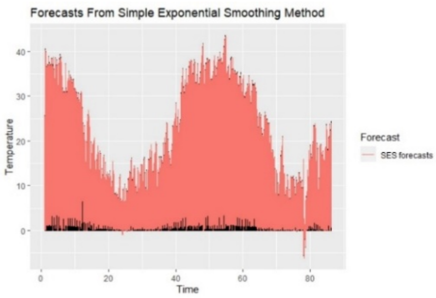

Figure 21: SES temperature forecasting model

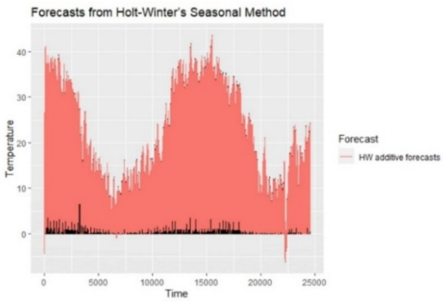

Figure 22: HWES temperature forecasting model

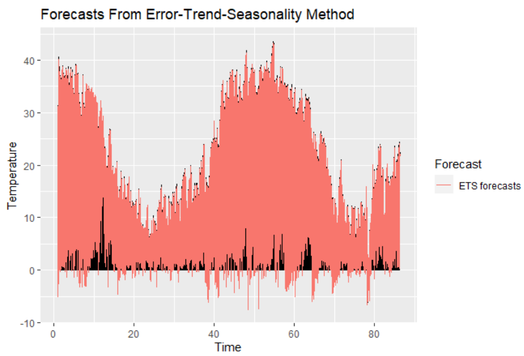

Figure 20: ETS temperature forecasting model

4.6.1 Evaluation of Performance Ratio

Table 9 represents the comparison of the actual PR and TPR results and the predicted PR and TPR calculations obtained from solar and temperature forecasting models with Random Tree, Holt and method. For purpose of verifying daily and 1-step ahead predictions, it is seen that actual PR and TPR results and the predicted results are close to each other.

In Table 10, it is represented that the evaluation metrics of 1 day ahead and 1 step ahead energy production forecasting models of the PV system which are developed using the best solar radiation and temperature prediction values obtained from Random Tree models.

Method	Time	Predicted PR (%)	Actual PR (%)	Predicted TPR (%)	Actual TPR (%)
Random Tree	Daily	36.03	28.68	35.50	30.00
Random Tree	1-Step ahead	35.08	28.68	34.64	30.00
Damped Holt's Method	Daily	35.54	28.68	35.05	30.00
ARIMA	Daily	35.79	28.68	35.29	30.00

Table 9: PR and TPR results of solar radiation and temperature forecasting models

Method	RMSE	MAE	MAPE	MSE	DAC
Random Tree 1-Step ahead	21.74	9.32	41.31	472.64	83.19
Random Tree Daily	21.26	8.33	41.95	451.94	84.08

Table 10: Evaluation of forecasting energy production using predicted solar radiation and temperature values

5 Conclusion

Forecasting solar radiation and temperature through machine learning and time series analysis methods is a crucial subject in renewable energy. Precise predictions are essential for optimizing solar energy performance and integrating systems into the energy grid. Prediction accuracy is critical for the control and management of photovoltaic (PV) systems.

Currently, researchers have addressed solar radiation and temperature forecasting obtained from meteorological data in various studies. However, for the development of precise PV forecasting models, calculations should incorporate both solar radiation and temperature forecasts. This study suggests that accessing solar radiation and temperature forecasts through a single application would be of immense benefit to users.

The models for predicting solar radiation and temperature were created using several methods such as IBK, linear regression, MLP Regressor, random forest, random tree, and ML Perceptron. Standard statistical error indicators, including RMSE, MAPE, MABE, MAE, MSE, and directional accuracy, were used to evaluate the models' accuracy and precision performance. Comparing performance, the study observed that Random Tree classifier yielded the most accurate predictions for solar radiation and temperature. Time series analysis incorporating ARIMA, SES, HOLT, HOLT, HWES and ETS were implemented to forecast solar radiation and temperature. Results indicated that the Damped HOLT and ARIMA methods exceeded the other forecasting models, based on all error indicators assessed in the study. The performance of a PV system is assessed using parameters known as PR and TPR. This study involves computational determination and comparison of PR and TPR values pertaining to the top predicted models to the actual ones.

For the purpose of achieving optimal performance ratio, our future studies aim to develop hybrid solar radiation and temperature models using various techniques alongside current methods. These models will leverage big data analytics to improve predictive accuracy. The use of deep learning models, conventional learning models, and machine learning algorithms are proposed. Optimization techniques and the incorporation of additional variables, such as air pollutant data, are also being explored. These advancements in solar radiation modeling will contribute to achieving optimal performance ratios in PV energy systems.

References

[Abdullah et al., 2020] Abdullah, H. M., Kamel, R. M., El-Sayed, M.: 'Non-communication and artificial neural network based photovoltaic monitoring using the existing impedance relay'; Sustainable Energy, Grids and Networks, 22 (2020), 100335. https://doi.org/10.1016/j.segan.2020.100335

[Ahmad et al., 2018] Ahmad, M. W., Mourshed, M., Rezgui, Y.: 'Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression'; Energy, 164 (2018), 465–474. https://doi.org/10.1016/j.energy.2018.08.207

[AKCAN et al., 2020] AKCAN, E., KUNCAN, M., MİNAZ, M. R.: 'PVsyst Yazılımı İle 30 Kw Şebekeye Bağlı Fotovoltaik Sistemin Modellenmesi ve Simülasyonu'; European Journal of Science and Technology, , 18 (2020), 248–261. https://doi.org/10.31590/ejosat.685909

[Aldous, 1991] Aldous, D.: 'The Continuum Random Tree. I'; The Annals of Probability (1991). https://doi.org/10.1214/aop/1176990534

[Alsharif et al., 2019] Alsharif, M. H., Younes, M. K., Kim, J.: 'Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea'; Symmetry, 11, 2 (2019), 1–17. https://doi.org/10.3390/sym11020240

[Andrade and Bessa, 2017] Andrade, J. R., Bessa, R. J.: 'Improving Renewable Energy Forecasting with a Grid of Numerical Weather Predictions'; IEEE Transactions on Sustainable Energy, 8, 4 (2017), 1571–1580. https://doi.org/10.1109/TSTE.2017.2694340

[Antonanzas et al., 2016] Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F. J., Antonanzas-Torres, F.: 'Review of photovoltaic power forecasting'; Solar Energy, 136 (2016), 78–111. https://doi.org/10.1016/j.solener.2016.06.069

[Azadeh et al., 2013] Azadeh, A., Babazadeh, R., Asadzadeh, S. M.: 'Optimum estimation and forecasting of renewable energy consumption by artificial neural networks'; Renewable and Sustainable Energy Reviews, 27 (2013), 605–612. https://doi.org/10.1016/j.rser.2013.07.007

[Azimi et al., 2016] Azimi, R., Ghayekhloo, M., Ghofrani, M.: 'A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting'; Energy Conversion and Management (2016). https://doi.org/10.1016/j.enconman.2016.04.009

[Azmi et al., 2023] Azmi, M. S. F. M., Hussain, M. H., Rahim, S. R. A., Mid, E. C., Shaari, A. S., Hashim, N., et al.: 'Hybrid Cooling System for Solar Photovoltaic Panel'; Journal of Physics: Conference Series, 2550, 1 (2023). https://doi.org/10.1088/1742-6596/2550/1/012004

[Bashir et al., 2021] Bashir, N., Irwin, D., Shenoy, P.: 'A Probabilistic Approach to Committing Solar Energy in Day-ahead Electricity Markets'; Sustainable Computing: Informatics and Systems (2021). https://doi.org/10.1016/j.suscom.2020.100477

[Bracale et al., 2013] Bracale, A., Caramia, P., Carpinelli, G., Di Fazio, A. R., Ferruzzi, G.: 'A Bayesian method for Short-Term probabilistic forecasting of photovoltaic generation in smart grid operation and control'; Energies, 6, 2 (2013), 733–747. https://doi.org/10.3390/en6020733

[Breiman, 2001] Breiman, L.: 'Random forests'; Machine Learning (2001). https://doi.org/10.1023/A:1010933404324

[BUDAK, 2018] BUDAK, H.: 'Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım'; Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 22, Özel (2018), 21. https://doi.org/10.19113/sdufbed.01653

[Choo and Wei, 2022] Choo, Y. M., Wei, W.: 'Salt hydrates as phase change materials for photovoltaics thermal management'; Energy Science and Engineering, 10, 5 (2022), 1630–1642. https://doi.org/10.1002/ese3.1007

[Chu et al., 2015] Chu, Y., Urquhart, B., Gohari, S. M. I., Pedro, H. T. C., Kleissl, J., Coimbra, C. F. M.: 'Short-term reforecasting of power output from a 48 MWe solar PV plant'; Solar Energy, 112 (2015), 68–77. https://doi.org/10.1016/j.solener.2014.11.017

[Çiftçi et al., 2020] Çiftçi, S., Solak, M., Kuncan, M.: 'Powered by the sun: designing and analyzing technical and economic aspects of a school sustained by photovoltaics'; Journal of Mechatronics and Artificial Intelligence in Engineering, 1, 1 (2020), 21–32. https://doi.org/10.21595/jmai.2020.21499

[Cornejo-Bueno et al., 2019] Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S.: 'Machine learning regressors for solar radiation estimation from satellite data'; Solar Energy, 183, March (2019), 768–775. https://doi.org/10.1016/j.solener.2019.03.079

[Cotfas, D., Marzband, M., Cotfas, P., Siroux, M., & Sera, 2022] Cotfas, D., Marzband, M., Cotfas, P., Siroux, M., & Sera, D.: 'Editorial: forecasting solar radiation, photovoltaic power and thermal energy production applications'; Frontiers in Energy Research, 10 (2022). https://doi.org/https://doi.org/10.3389/fenrg.2022.1115096

[Das et al., 2018] Das, U. K., Tey, K. S., Seyedmahmoudian, M., Mekhilef, S., Idris, M. Y. I., Van Deventer, W., et al.: 'Forecasting of photovoltaic power generation and model optimization: A review'; Renewable and Sustainable Energy Reviews, 81, April 2017 (2018), 912–928. https://doi.org/10.1016/j.rser.2017.08.017

[Diagne et al., 2013] Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N.: 'Review of solar irradiance forecasting methods and a proposition for small-scale insular grids'; Renewable and Sustainable Energy Reviews, 27 (2013), 65–76. https://doi.org/10.1016/j.rser.2013.06.042

[Dierauf et al., 2013] Dierauf, T., Growitz, A., Kurtz, S., Hansen, C.: 'Weather-Corrected Performance Ratio'; NREL Technical Report NREL/TP-5200-57991, , April (2013), 1–16.

[Ding et al., 2011] Ding, M., Wang, L., Bi, R.: 'An ANN-based approach for forecasting the power output of photovoltaic system'; Procedia Environmental Sciences, 11, PART C (2011), 1308–1315. https://doi.org/10.1016/j.proenv.2011.12.196

[Fara et al., 2021] Fara, L., Diaconu, A., Craciunescu, D., Fara, S.: 'Forecasting of Energy Production for Photovoltaic Systems Based on ARIMA and ANN Advanced Models'; International Journal of Photoenergy, 2021 (2021). https://doi.org/10.1155/2021/6777488

[Gaboitaolelwe et al., 2023] Gaboitaolelwe, J., Zungeru, A. M., Yahya, A., Lebekwe, C. K., Vinod, D. N., Salau, A. O.: 'Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison'; IEEE Access, 11, March (2023), 40820–40845. https://doi.org/10.1109/ACCESS.2023.3270041

[Goswami and Sadhu, 2021] Goswami, A., Sadhu, P. K.: 'Degradation analysis and the impacts on feasibility study of floating solar photovoltaic systems'; Sustainable Energy, Grids and Networks, 26 (2021), 100425. https://doi.org/10.1016/j.segan.2020.100425

[Hall, 1999] Hall, M. A.: 'Correlation-based Feature Selection for Machine Learning';, , April (1999).

[Hall et al., 2009] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H.: 'The WEKA data mining software'; ACM SIGKDD Explorations Newsletter (2009). https://doi.org/10.1145/1656274.1656278

[Ihaka and Gentleman, 1996] Ihaka, R., Gentleman, R.: 'R: A Language for Data Analysis and Graphics'; Journal of Computational and Graphical Statistics (1996). https://doi.org/10.1080/10618600.1996.10474713

[Inman et al., 2013] Inman, R. H., Pedro, H. T. C., Coimbra, C. F. M.: 'Solar forecasting methods for renewable energy integration'; Progress in Energy and Combustion Science, 39, 6 (2013), 535–576. https://doi.org/10.1016/j.pecs.2013.06.002

[International Electrotechnical Commission, 1998] International Electrotechnical Commission: 'Photovoltaic system performance monitoring-guidelines for measurement, data exchange and analysis'; IEC 61724 (1998).

[Ji and Chee, 2011] Ji, W., Chee, K. C.: 'Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN'; Solar Energy, 85, 5 (2011), 808–817. https://doi.org/10.1016/j.solener.2011.01.013

[Jiawei Han, 2012] Jiawei Han, M. K. and J. P.: 'Data Mining: Concepts and Techniques, Third Edition - Books24x7'; (2012).

[Kapur et al., 2017] Kapur, B., Ahluwalia, N., R, S.: 'Comparative Study on Marks Prediction using Data Mining and Classification Algorithms'; International Journal of Advanced Research in Computer Science, 8, 3 (2017), 632–636. https://doi.org/10.26483/IJARCS.V8I3.3066

[Kardakos et al., 2013] Kardakos, E. G., Alexiadis, M. C., Vagropoulos, S. I., Simoglou, C. K., Biskas, P. N., Bakirtzis, A. G.: 'Application of time series and artificial neural network models in short-term forecasting of PV power generation'; In Proceedings of the Universities Power Engineering Conference (2013). https://doi.org/10.1109/UPEC.2013.6714975

[Kim et al., 2019] Kim, S. G., Jung, J. Y., Sim, M. K.: 'A two-step approach to solar power generation prediction based on weather data using machine learning'; Sustainability (Switzerland), 11, 5 (2019). https://doi.org/10.3390/SU11051501

[Kırbaş, 2018] Kırbaş, İ.: 'İstatistiksel Metotlar ve Yapay Sinir Ağları Kullanarak Kısa Dönem Çok Adımlı Rüzgâr Hızı Tahmini'; SAÜ Fen Bilimleri Enstitüsü Dergisi, , January (2018), 1–1. https://doi.org/10.16984/saufenbilder.305224

[Leva et al., 2017] Leva, S., Dolara, A., Grimaccia, F., Mussetta, M., Ogliari, E.: 'Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power'; Mathematics and Computers in Simulation, 131 (2017), 88–100. https://doi.org/10.1016/j.matcom.2015.05.010

[Li et al., 2020] Li, G., Duan, Z., Liang, L., Zhu, H., Hu, A., Cui, Q., et al.: 'Outlier data mining method considering the output distribution characteristics for photovoltaic arrays and its application'; Energy Reports, 6 (2020), 2345–2357. https://doi.org/10.1016/j.egyr.2020.08.034

[Libra et al., 2021] Libra, M., Petrik, T., Poulek, V., Tyukhov, I. I., Kourim, P.: 'Changes in the Efficiency of Photovoltaic Energy Conversion in Temperature Range with Extreme Limits'; IEEE Journal of Photovoltaics, 11, 6 (2021), 1479–1484. https://doi.org/10.1109/JPHOTOV.2021.3108484

[Liu et al., 2019] Liu, L., Zhan, M., Bai, Y.: 'A recursive ensemble model for forecasting the power output of photovoltaic systems'; Solar Energy, 189, July (2019), 291–298. https://doi.org/10.1016/j.solener.2019.07.061

[Loganathan et al., 2021] Loganathan, A. K., Stonier, A. A., Maheswari, Y. U.: 'Performance enhancement of a photovoltaic module using solar functional coatings'; Journal of Materials Science: Materials in Electronics, 32, 1 (2021), 1242–1257. https://doi.org/10.1007/s10854-020-04897-0

[Mathiesen et al., 2013] Mathiesen, P., Collier, C., Kleissl, J.: 'A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting'; Solar Energy, 92 (2013), 47–61. https://doi.org/10.1016/j.solener.2013.02.018

[Mathiesen and Kleissl, 2011] Mathiesen, P., Kleissl, J.: 'Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States'; Solar Energy, 85, 5 (2011), 967–977. https://doi.org/10.1016/j.solener.2011.02.013

[Mellit and Pavan, 2010] Mellit, A., Pavan, A. M.: 'A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy'; Solar Energy, 84, 5 (2010), 807–821. https://doi.org/10.1016/j.solener.2010.02.006

[Natarajan et al., 2021] Natarajan, Y., Kannan, S., Selvaraj, C., Mohanty, S. N.: 'Forecasting energy generation in large photovoltaic plants using radial belief neural network'; Sustainable Computing: Informatics and Systems, 31, May (2021), 100578. https://doi.org/10.1016/j.suscom.2021.100578

[Ogliari et al., 2017] Ogliari, E., Dolara, A., Manzolini, G., Leva, S.: 'Physical and hybrid methods comparison for the day ahead PV output power forecast'; Renewable Energy, 113 (2017), 11–21. https://doi.org/10.1016/j.renene.2017.05.063

[Oğuzlar, 2003] Oğuzlar, A.: 'Veri Ön İşleme'; Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, , 21 (2003). https://doi.org/10.18070/euiibfd.90046

[Olive, 2017] Olive, D. J.: 'Linear regression'; Linear Regression (2017). https://doi.org/10.1007/978-3-319-55252-1

[Pelland et al., 2013] Pelland, S., Remund, J., Kleissl, J., Oozeki, T., De Brabandere, K.: 'Photovoltaic and Solar Forecasting : State of the Art'; (2013), 40.

[Rahman et al., 2023] Rahman, Y. A., Amin, N., Masarrang, M., Pirade, Y., Yunfa, N. P., Santoso, R.: 'Monitoring PV Output Power in Tropical Climates under Surface Temperature Distribution Area'; IOP Conference Series: Earth and Environmental Science, 1157, 1 (2023). https://doi.org/10.1088/1755-1315/1157/1/012037

[Ramsami and Oree, 2015] Ramsami, P., Oree, V.: 'A hybrid method for forecasting the energy output of photovoltaic systems'; Energy Conversion and Management, 95 (2015), 406–413. https://doi.org/10.1016/j.enconman.2015.02.052

[Raza et al., 2016] Raza, M. Q., Nadarajah, M., Ekanayake, C.: 'On recent advances in PV output power forecast'; Solar Energy, 136 (2016), 125–144. https://doi.org/10.1016/j.solener.2016.06.073

[Sengupta et al., 2015] Sengupta, M., Habte, A., Kurtz, S., Dobos, A., Wilbert, S., Lorenz, E., et al.: 'Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications';, , NREL/TP-5D00-63112 (2015), 251 pages. Retrieved from http://www.nrel.gov/docs/fy15osti/63112.pdf

[Sobri et al., 2018] Sobri, S., Koohi-Kamali, S., Rahim, N. A.: 'Solar photovoltaic generation forecasting methods: A review'; Energy Conversion and Management, 156, May 2017 (2018), 459–497. https://doi.org/10.1016/j.enconman.2017.11.019

[Viscondi and Alves-Souza, 2021] Viscondi, G. de F., Alves-Souza, S. N.: 'Solar irradiance prediction with machine learning algorithms: A Brazilian case study on photovoltaic electricity generation'; Energies, 14, 18 (2021). https://doi.org/10.3390/en14185657

[Xiang et al., 2021] Xiang, X., Sun, Y., Deng, X.: 'Short Time Solar Power Forecasting Using Persistence Extreme Learning Machine Approach'; E3S Web of Conferences, 294 (2021). https://doi.org/10.1051/e3sconf/202129401002

[Yang et al., 2012] Yang, D., Jirutitijaroen, P., Walsh, W. M.: 'Hourly solar irradiance time series forecasting using cloud cover index'; Solar Energy, 86, 12 (2012), 3531–3543. https://doi.org/10.1016/j.solener.2012.07.029

[Yang et al., 2015] Yang, D., Sharma, V., Ye, Z., Lim, L. I., Zhao, L., Aryaputera, A. W.: 'Forecasting of global horizontal irradiance by exponential smoothing, using decompositions'; Energy, 81 (2015), 111–119. https://doi.org/10.1016/j.energy.2014.11.082

[Yang and Yang, 2014] Yang, Z. R., Yang, Z.: 'Artificial Neural Networks'; In Comprehensive Biomedical Physics (2014). https://doi.org/10.1016/B978-0-444-53632-7.01101-1

[Zafarani et al., 2018] Zafarani, R., Eftekharnejad, S., Patel, U.: 'Assessing the Utility of Weather Data for Photovoltaic Power Prediction'; (2018). Retrieved from http://arxiv.org/abs/1802.03913

[Zendehboudi et al., 2018] Zendehboudi, A., Baseer, M. A., Saidur, R.: 'Application of support vector machine models for forecasting solar and wind energy resources: A review'; Journal of Cleaner Production, 199 (2018), 272–285. https://doi.org/10.1016/j.jclepro.2018.07.164

[Zhang et al., 2014] Zhang, Y., Beaudin, M., Zareipour, H., Wood, D.: 'Forecasting Solar Photovoltaic power production at the aggregated system level'; In 2014 North American Power Symposium, NAPS 2014 (2014). https://doi.org/10.1109/NAPS.2014.6965389

[Zheng et al., 2011] Zheng, Z. W., Chen, Y. Y., Huo, M. M., Zhao, B.: 'An overview: The development of prediction technology of wind and photovoltaic power generation'; Energy Procedia, 12 (2011), 601–608. https://doi.org/10.1016/j.egypro.2011.10.081