

# Global solar radiation forecast using an ensemble learning approach

Debani Prasad Mishra<sup>1</sup>, Subhrajit Jena<sup>1</sup>, Rudranarayan Senapati<sup>2</sup>, Atman Panigrahi<sup>1</sup>,  
Surender Reddy Salkuti<sup>3</sup>

<sup>1</sup>Department of Electrical and Electronics Engineering, IIIT Bhubaneswar, Odisha, India

<sup>2</sup>Department Electrical Engineering, KIIT Deemed to be University, Odisha, India

<sup>3</sup>Department of Railroad and Electrical Engineering, Woosong University, Daejeon, Republic of Korea

## Article Info

### Article history:

Received Sep 8, 2022

Revised Nov 7, 2022

Accepted Nov 25, 2022

### Keywords:

Electricity grid

Ensemble methods

Extreme gradient boosting

Random forest

Solar energy

Sustainable utilization

## ABSTRACT

With the increase in demand for solar power, a solar power forecasting model is of maximum importance to allow a higher level of integration of non-conventional energy into the existing electricity grid. With the advancement in data availability, there's a good time to use data-driven algorithms for enhanced prediction of solar energy generation. Gathering and analyzing data can predict solar energy generation and mitigate the impact of solar intermittency. During this research, we explore automatically creating prediction models that are site-specific utilizing machine learning to generate solar radiation from meteorological station weather forecast reports, and from the predicted solar radiation corresponding solar power output can be calculated depending upon the characteristics of the solar PV system used. The challenge is to enhance the accuracy of the forecast. Ensemble techniques like random forest (RF) and extreme gradient boosting (XGBoost) are well suited for solar radiation prediction as they improve stability as well as combine several machine learning models to reduce variation and bias which outperforms the majority of models, as a result making them a perfect model in the field of solar energy prediction.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Surender Reddy Salkuti

Department of Railroad and Electrical Engineering, Woosong University

Jayang-dong, Dong-gu, Daejeon-34606, Republic of Korea

Email: surender@wsu.ac.kr

## NOMENCLATURE

|         |   |
|---------|---|
| ANN     | : Artificial neural network                           |
| ARIMA   | : Autoregressive integrated moving average            |
| CRO-ELM | : Coral reefs optimization-extreme learning machine   |
| ELM     | : Extreme learning machine                            |
| GANN    | : Genetic algorithm neural network                    |
| GGA-ELM | : Grouping genetic algorithm-extreme learning machine |
| GRNN    | : Generalised regression neural network               |
| KELM    | : Kernel extreme learning machine                     |
| MABE    | : Mean absolute bias error                            |
| MAE     | : Mean absolute error                                 |
| MAPE    | : Mean absolute percentage error                      |
| MBE     | : Mean bias error                                     |

|           |   |
|-----------|---|
| MSE       | : Mean square error   |
| R2        | : R2 score  |
| RF        | : Random forest   |
| RMSE      | : Root mean square error  |
| RRMSE     | : Robust root mean square error   |
| SOM-OPELM | : Self-organizing feature map–optimally pruned extreme learning machine |
| SVM       | : Support vector machine  |

## 1. INTRODUCTION

The primary objective of utilizing non-conventional energy systems is to mitigate global climate change, provide more access to energy and improve energy security [1]. By using sustainable energy and guaranteeing that all residents have access to inexpensive, dependable, sustainable, and contemporary energy, sustainable development can be made possible [2]–[4]. Considering the stress and needs, solar power radiation can be an optimal solution for non-conventional energy sources [5]. The challenges and limitations of solar power are reduced significantly due to advancements in photovoltaic technologies which in turn increases the efficiency of energy conversion and reduces panel installation and electricity cost notably. Solar power is the future, considering the very fact that it is an inexhaustible energy source and requires low installation costs [6]. PV panels often cannot provide stable electric power output due to variations in weather conditions, the facility grid stability is reduced significantly while integrating the photovoltaic power into the power grid [7]. As a stable power system is a necessity, a particular forecasting technique is required for the stable and safe integration of the photovoltaic power into the power grid [8]. In addition to helping to stabilize the grid, a precise forecasting model technique is essential for managing storage, developing an energy road map for congestion management, and estimating reserves. The availability of data has made it possible to use deep learning and machine learning techniques [9].

In order to manage energy use optimally in the present, safely operate power systems, and balance consumption and production, the predictive analysis will be crucial [10]. The goal of the vision is to introduce and use the most recent technology to create electrical networks that are more secure, effective, eco-friendly, and reliable [11]. With the advancement in accurate forecasting of meteorological and hydrological variables like precipitation, evaporation, and temperature humidity has made it possible for predicting solar energy generation in a more efficient manner [11]. PV prediction models can be used by consumers to coordinate their use with on-site power generation and, as a result, maximize their profitability. One of the key advantages of data-driven models is that they utilize less time to make judgments on power system planning and take less time to perform predictions [12]. Accurate prediction of energy produced by PV systems has been identified together with the major challenges as it allows grid operators to manage electricity generation by making informed decisions which in turn reduces the uncertainties and cost. Several forecasting methods presented in the literature are described in Table 1.

Table 1. Several forecasting methods earlier proposed by researchers

|      | Best Predictive models | Alternative models | Input variables   | Study area | Performance indices         | Time scale |
|------|------------------------|--------------------|---|------------|-----------------------------|------------|
| [13] | ELM                    | ANN                | Altitudes, latitudes, longitudes, land-surface temperatures   | Turkey     | R2, MBE, RMSE               | Daily      |
| [14] | ELM                    | SVM                | Ozone, cloudiness   | Spain      | R2, MSE                     | Daily      |
| [15] | CRO-ELM                | SVM                | Ozone, water vapor  | Spain      | R2, MAE, RMSE, Bias         | Daily      |
| [16] | ELM                    | ARIMA, SVM         | Air temperature, relative humidity  | Italy      | Bias                        | Daily      |
| [17] | KELM                   | SVM                | Maximum and minimum air temperature   | Iran       | MABE, RMSE, RRMSE           | Daily      |
| [18] | ELM                    | SVM, ANN           | Sunshine duration, Air temperature  | Iran       | R2, MABE, MAPE, RMSE, RRMSE | Daily      |
| [19] | GGA-ELM                | ELM                | Long-wave radiation, long-wave flux, short-wave flux, wind velocity, cloud fraction, water vapor, air temperature | Spain      | RRMSE                       | Hourly     |
| [20] | SOM-OPELM              | ARIMA              | Sunshine duration   | USA        | MAPE, MAE, MBE, RMSE        | Hourly     |
| [21] | ELM, GANN              | GRNN, RF           | Sunshine duration   | China      | MAE, RMSE                   | Daily      |

One of the most predominantly used machine learning techniques, the support vector machine has been used to a great extent in building energy and for predictions of non-conventional energy applications. Even with a small sample of datasets, the approach is incredibly efficacious for resolving non-linear problems. In order to reduce the generalization error's upper bound, which is made up of the sum of the training error and a confidence level, the support vector machine utilizes the structural risk minimization (SRM) concept [22]. Applying the fundamental principles of Support Vector Machine to regression problems entails adding a kernel function, non-linearly transforming the input space into a higher-dimensional feature space, and then performing a linear regression in this feature space [23], [24]. But still, this technique has promised researchers over the years over certain datasets. It was discovered that the suggested technique outperformed the ANN in terms of performance. Shi *et al.* [25] also suggested a support vector machine and weather classification-based PV forecast approach. The outcomes demonstrated that the suggested prediction model for grid-connected systems was successful and promising. Utilizing support vector machines, Yousif and Kazem research [26] developed solar photovoltaic power output. The proposed model projected photovoltaic current using inputs such as solar radiation and ambient air temperature. Kazem and Yousif [27] employed a support vector machine model and evaluated its performance in comparison to multi-layer perceptron, and generalized feed-forward networks (GFF).

In artificial intelligence (AI) and machine learning, deep learning models are regarded as a new paradigm of learning. Recent years have seen a substantial increase in interest in deep learning due to its ability to handle complex data. Various architectures of ANN's from simple ANN networks to more complex models like an auto-encoder [28] and Long short-term memory (LSTM) [29] networks have been effectively and successfully utilized to forecast renewable energy. A method based on artificial neural networks was developed by Hiyama and Kitabayashi [30] to predict the maximum power output from a PV system. The author's input features included solar radiation, wind speed, and outside air temperature [31]. A hybrid multi-layer feed-forward neural network was created by Sulaiman *et al.* [32] to estimate the output from a grid-connected PV system.

The published works show that many approaches have been used to predict solar power or solar radiation output. Amongst the employed techniques artificial neural network is one of the favored techniques. However, an artificial neural network requires the user to provide several model parameters, for instance, the number of neurons in hidden layers, the number of hidden layers, and the number of training epochs. Two of the most prominent machine learning techniques support vector machines and artificial neural networks, showed instability issues [33]. Due to the instability, even slight changes in the input data could cause significant variances in the anticipated values. To overcome these instability issues, a more advanced machine learning algorithm, Ensemble Learning was developed.

Ensemble learning is a machine learning technique that involves training many base learners and combining their output to address a single problem. The fundamental tenet is that the aggregate output of the weak learners or the base learners should generally be more accurate than the output of any one learner. Very little research has been done on ensemble techniques for predicting solar radiation, which indicates that ensemble-based methods like RF and gradient boosting and extreme gradient boosting techniques have not been extensively studied. Considering that they are able to overcome the shortcomings, ensemble-based strategies typically outperform individual learners who build them [25]. The approach of ensemble techniques has drawn a lot of interest and is now in demand across various industries. Utilizing ensemble-based methods with solar photovoltaic systems motivates because the majority of the earlier research works are centered around regressive methods, support vector machines, and artificial neural networks and their variants, and these ensemble-based algorithms are more computationally efficient in comparison to the other widely used algorithms. The paper performs the research on the performances of two major Ensemble technique methods i.e., RF and XGBoost with Hyper-parameter tuning in solar radiation prediction from future weather data produced by the meteorological station.

## 2. THEORY OF FORECASTING MODEL

### 2.1. Random forest (RF)

RF is an ensemble-based bagging machine learning algorithm comprising a significant number of decision trees. Decision trees are used as the base learners or the weak learners in the RF models. The working of a RF is depicted in Figure 1. In RF, the performances of individual base learners i.e., the decision trees are boosted by the aggregation of individual tree results. The main trademark of RF is random feature sampling and random row sampling while selecting a set of rows and features from a dataset to train a particular decision tree. Cross-validation is not necessary when using RF because they can do out-of-bag error estimation as part of the forest-building process. By randomly selecting data from the initial training dataset with replacement, RF initially creates numerous additional training data sets. The size of the new

training dataset is the same as the previous ones, however, sampling with replacement may result in some observations being duplicated [34]. Each decision tree model has high variance as gets trained with particular samples but as the final output depends on the aggregation of all the individual model's output, the final output has low variance.

**2.2. XGBoost**

XGBoost is an optimized distributed gradient boosting ensemble machine learning algorithm that uses a gradient boosting framework. The working of XGBoost is depicted in Figure 2. At the University of Washington, a research endeavor led to the creation of the XGBoost algorithm which led to a major advancement in the Machine Learning domain. XGBoost also uses Decision trees as its base for weak learners. In XGBoost, the Decision trees are built in a sequential manner. The weak learners are trained sequentially and each weak learners are, therefore dependent on each other [35]. In XGBoost, weight plays a significant role. Before being fed into the decision tree that predicts the results, each independent variable is given a weight. The weight of the variables that were incorrectly predicted by the base or the weak learners are increased and after that, they are fed to the next subsequent learner. These distinct weak learners are then combined to produce a model that is more precise and accurate. The high execution speed out of the core computation of XGBoost makes it a favorite among data scientists [36].

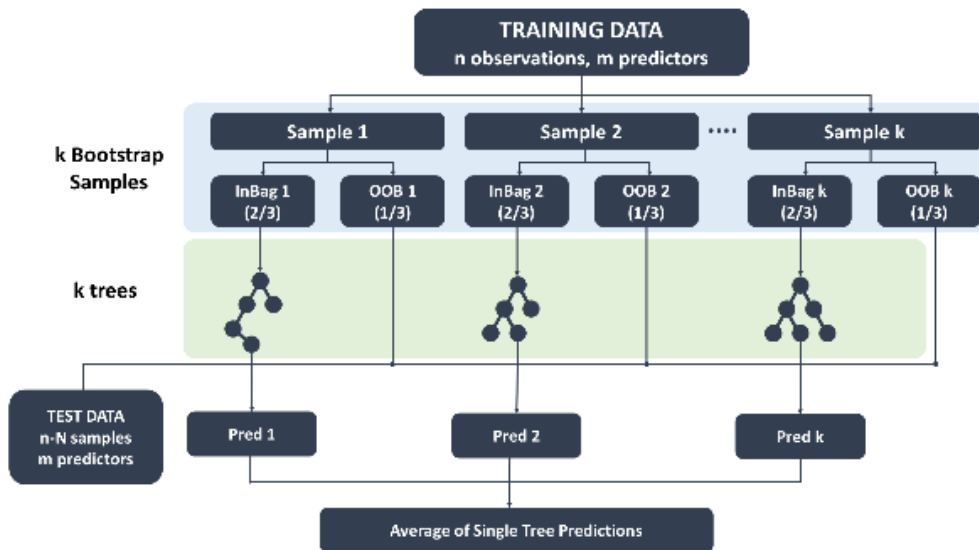


Figure 1. Working of RF

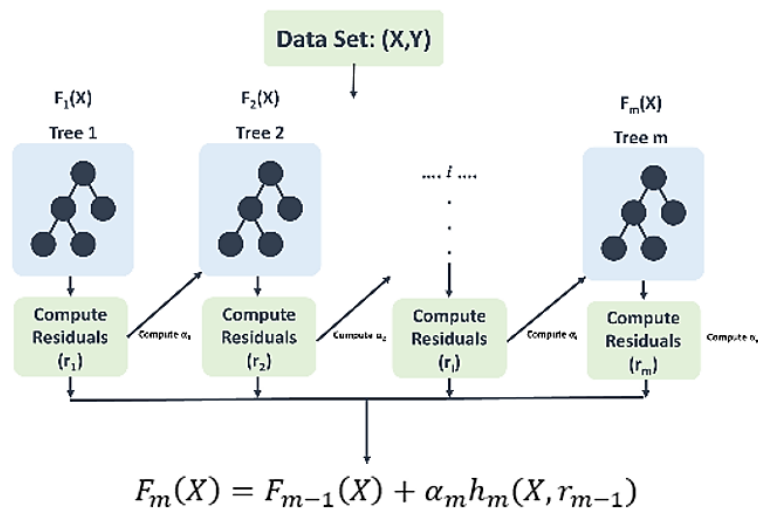


Figure 2. Working of XGBoost

### 3. METHODOLOGY

#### 3.1. Data Description

These datasets are meteorological data from the Hawaii Space Exploration Analog and Simulation (HI-SEAS) weather station between mission IV and mission V on which the model is trained and tested to get the best result possible. Weather parameters in the dataset include temperature, humidity, day time, wind speed, sunrise/sunset time, wind direction, and barometric pressure. The input dataset contains 15 min interval between each instance of weather parameters data. The model predicts the Solar Radiation as output. The model output calculated further with characteristic parameters of PV panel used gives the Solar Power output. The everyday hourly values of radiation are shown in Figure 3 and Figure 4 represents the combination of pressure and temperature for different values of radiation.

Figure 5 represents the correlation plot between all the weather parameters. The correlation matrix represents the correlation coefficient between two variables which describes the extent of the linear relationship between them. The diagonal elements of correlation matrix will have value 1 because it is the cross-section of same weather parameters. The values near to 0 indicate that the features are very minimally related to each other while the values near 1 and -1 indicated the features are maximally related to each other.

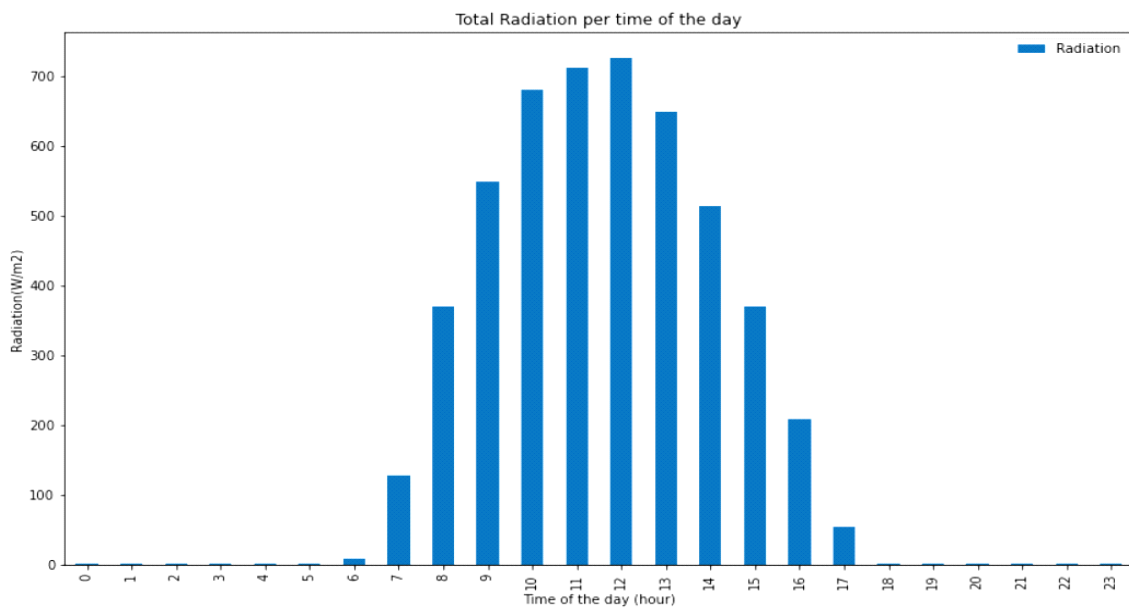


Figure 3. Everyday hourly values of radiation

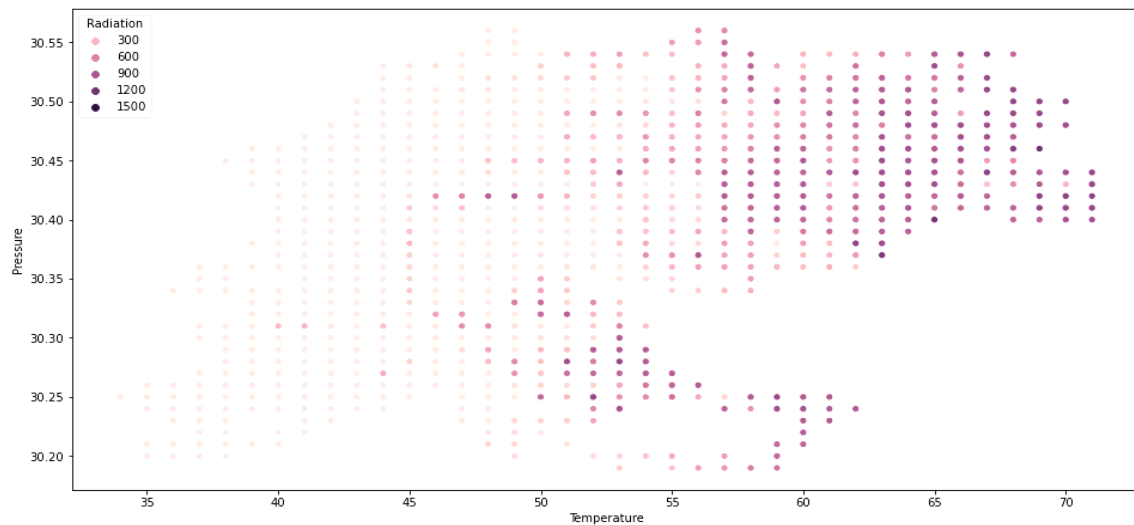


Figure 4. Pressure and temperature for different value of radiation

### 3.2. Evaluation indices

#### 3.2.1. Root mean square error (RMSE)

The standard deviation of the residuals is known as the root mean square error (RMSE). The residual is a fraction of the distance from the fallback line, which is the information hotspot. RMSE is the percentage of how much these residuals are fanned out. At the end of the day, you'll see how the information might be best suited. Mean squared error is commonly used to validate experimental results in climatology, estimation, and multivariate studies. The formula says:

$$RMSE = \sqrt{\overline{(f - o)^2}} \quad (1)$$

where  $f$  is predicted values (model output),  $o$  is actual values. The mean is indicated by the bar above the squared differences. The slightly different notation can be used to write the same formula as follows:

$$RMSE = \left[ \frac{\sum_{i=1}^N (Z_{fi} - Z_{oi})^2}{N} \right]^{\frac{1}{2}} \quad (2)$$

where,  $(Z_{fi} - Z_{oi})^2$  is differences, squared, and  $N$  is the sample size.

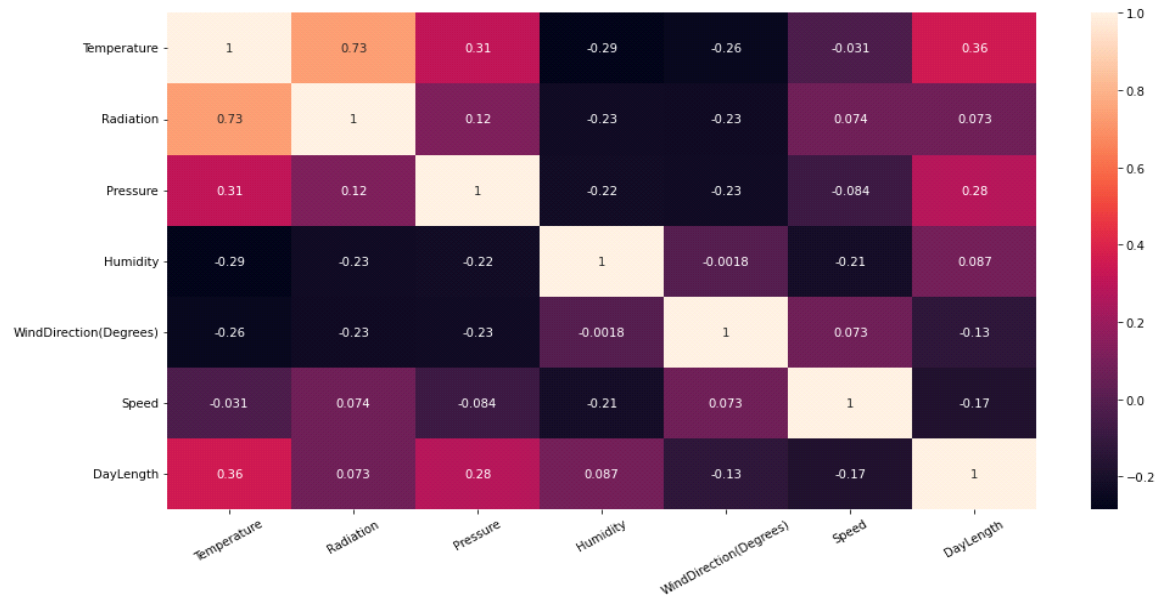


Figure 5. Correlation plot between all the weather parameters

#### 3.2.2. R<sup>2</sup> score

The coefficient of determination, also known as the R<sup>2</sup> score, is used to evaluate the regressive model's accuracy. It operates by calculating the variation in the predictions that the dataset can explain. It is used to determine how accurately the model predicts the observed results based on the ratio of the total deviations of the results it describes, and it is expressed as,

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (3)$$

where  $SS_{tot}$  denotes the total sum of errors and  $SS_{res}$  denotes the sum of squares of the residual errors.

### 3.3. Hyperparameter tuning

Hyperparameter tuning consists of finding a setup optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any dataset. That combination of hyper-parameters maximizes the model's performance and minimizes the predefined loss function to supply better accuracy. Hyper-parameters are specific to algorithms themselves, so we will calculate their values from the data. We use hyperparameters to calculate the model parameters.

### 3.3.1. Random search cross-validation

The most effective method for discovering the ideal collection of hyper-parameters for a machine learning model is random search. Using random draws from a specified set of hyper-parameter distributions, the randomized search meta-estimator algorithm trains and assesses a number of models. After training  $N$  distinct models with various randomly chosen hyper-parameter combinations, the algorithm chooses the best successful version of the model it has seen, giving you a model trained on a nearly ideal set of hyper-parameters.

### 3.3.2. Optuna

Optuna is a software framework for automating the optimization of hyper-parameters. By utilizing several samplers, including grid search, random, Bayesian, and evolutionary algorithms, it automatically determines the ideal hyper-parameter values. We can pass any Machine Learning algorithm as hyperparameters in Optuna and it will give the algorithm that gives the best result along with its hyperparameters.

## 4 MODEL TRAINING AND PREDICTION RESULTS

### 4.1. Random forest (RF) regression model

#### 4.1.1. Training and Hyperparameter tuning

The initial form of the data is  $x*y$  format, where  $x$  stands for the number of features and outputs and  $y$  for the overall number of instances. In this phase, the dataset is separated into training and testing sets. The effectiveness of the studied RF algorithms depends on the adjustment of hyperparameters, i.e., number of trees, number of features to consider at every split, minimum numbers of samples required to split a node, the maximum number of levels in the tree, minimum number of samples required at each leaf node and bootstrap (A dataset is randomly sampled with replacement using the statistical resampling approach known as “bootstrapping”). The Random search CV hyperparameter tuning gives the combination of the best set of parameters for a more accurate model.

#### 4.1.2. Prediction and result

The predictive performance of the RF regressor model is illustrated in Figure 6. The graph illustrates the plots of radiation values predicted by the RF model at different unix times vs measured values (actual value) in the testing dataset. The outcomes illustrate the level of a linear relationship and demonstrate how accurate the model can forecast solar radiation. At some unix time, larger discrepancies between real and anticipated values are seen due to a higher variation of solar radiation. In spite of that, the built RF model showed strong non-linear mapping generalizations ability and can be efficient in the prediction of solar radiation. The model was evaluated using  $R^2$  score evaluation indices which resulted to be 0.809 and RMSE score came out to be 108.55.

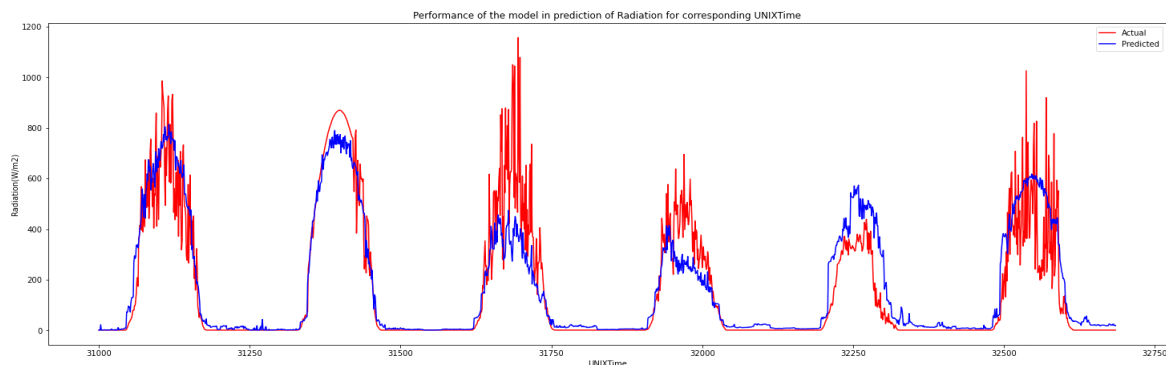


Figure 6. Radiation values predicted by RF model at different unix time vs measured value

### 4.2. XGBoost model

#### 4.2.1. Training and Hyperparameter tuning

The initial form of the data is  $x*y$  format, where  $x$  stands for the number of features and outputs and  $y$  for the overall number of instances. In this phase, the dataset is split into the training set, validation set, and testing set. While fine-tuning model hyperparameters, the validation set is used to provide an unbiased evaluation of a model fit on the training dataset.

The adjustment of hyperparameters increases the accuracy of the XGBoost model, i.e., learning rate (learning rate, simply refers to how quickly the model learns), early stopping (validation metric at least improve once in every round(s) to continue training), evals (it is a list of validation sets for which metrics will be evaluated during training), depth of the tree, 'num\_boost\_round' (number of trees to build). Optuna hyperparameter tuning is used to optimize the model parameters. During optimization, at each iteration, a new set of parameters is created and their loss value is evaluated. The set of parameters with less loss value is chosen as the best set of parameters. Then the model is built using those sets of parameters.

#### 4.2.2. Prediction and result

The predictive performance of the XGBoost model is illustrated in Figure 7. The graph illustrates the plots of radiation values predicted by the RF model at different Unix times vs measured values (actual value) in the testing dataset. The outcomes illustrate the level of a linear relationship and demonstrate how accurate the model can forecast solar radiation. The statistical evaluation indices,  $R^2$  and RMSE were used to appraise the model's performance as represented in Table 2. The XGBoost model is observed to show higher fluctuation in predicting solar radiation than the RF model. The model's  $R^2$  value came out to be 0.64 and RMSE score to be 122.12.

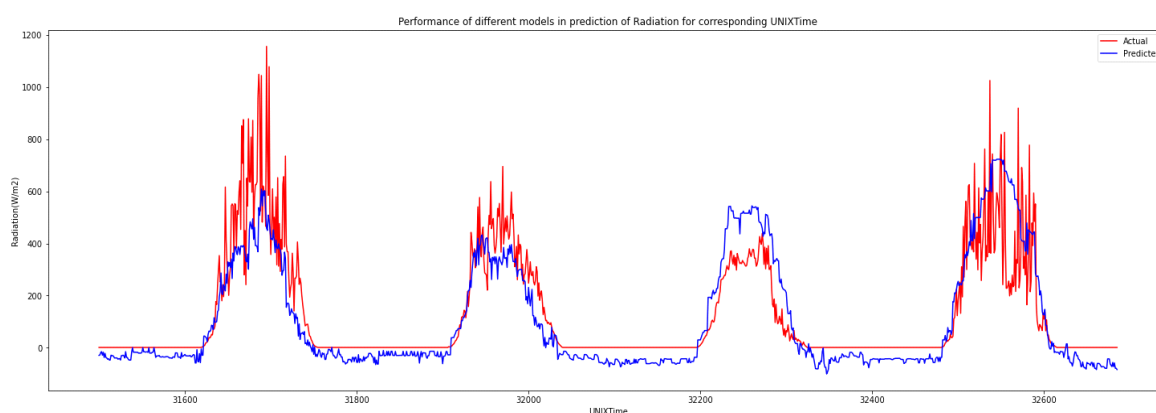


Figure 7. Radiation values predicted by RF model at different unix time vs measured value

Table 2. Statistical evaluation indices

| Models            | Performance Indices |          |
|-------------------|---------------------|----------|
|                   | $R^2$               | RMSE     |
| RF regressor      | 0.809494            | 108.55   |
| XGBoost regressor | 0.645419            | 122.1278 |

## 5. CONCLUSION

In this research work, the practicability of deploying tree-based ensemble methods (RF and XGBoost) to predict solar radiation which in turn evaluates the photovoltaic system power output. The capability of ensemble technique methods for predicting solar radiation has been verified with model prediction performances. Ensemble algorithms were shown to marginally outperform other popular machine learning techniques. The work also aimed to use tree-based ensemble methods to explain the significance of the input attributes. Based on several weather parameters, the developed machine learning models can be used to forecast solar radiation. Both RF (internal cross-validation) and XGBoost perform cross-validation and can be used to manage datasets with large dimensions. The modeling strategy is demonstrated a reliable one that can be used for real-time solar radiation prediction. There is still room for improvement which can lead to more accurate models. One of the future work directions is to extend the existing work for more generalized datasets.

## ACKNOWLEDGEMENTS

This research work was funded by "Woosong University's Academic Research Funding – 2022".

## REFERENCES

- [1] N. Lewis, "Powering the Planet," *MRS Bulletin*, vol. 32, no. 10, pp. 808–820, 2007, doi: 10.1557/mrs2007.168.







- [2] P. Nema, R. Nema, and S. Rangnekar, "A current and future state of art development of hybrid energy system using wind and PV-solar: A review," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 2096–2103, 2009, doi: 10.1016/j.rser.2008.10.006.
- [3] A. Zobaa, and R. Bansal, "Handbook of renewable energy technology," 2011, doi: 10.1142/7489.
- [4] D. Larcher, and J. Tarascon, "Towards greener and more sustainable batteries for electrical energy storage," *Nature Chemistry*, vol. 7, no. 1, pp. 19–29, 2014, doi: 10.1038/nchem.2085.
- [5] U. Economic and S. Council, "Progress towards the sustainable development goals", New York, NY, USA, Tech. Rep., 2017.
- [6] IEA, "Key world energy statistics 2020," Paris, 2020. [Online]. Available: [https://www.energiafed.be/sites/default/files/editor/Key\\_World\\_Energy\\_Statistics\\_2020\\_0.pdf](https://www.energiafed.be/sites/default/files/editor/Key_World_Energy_Statistics_2020_0.pdf)
- [7] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, "Regional PV power prediction for improved grid integration," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 7, pp. 757–771, 2010, doi: 10.1002/pip.1033.
- [8] M. Reking et al., "Connecting the Sun: Solar photovoltaics on the road to large-scale grid integration," Solar Business Hub, 2012. [Online]. Available: <https://resources.solarbusinesshub.com/solar-industry-reports/item/connecting-the-sun-solar-photovoltaics-on-the-road-to-large-scale-grid-integration>.
- [9] M. Zieher, M. Lange and U. Focken, "Variable renewable energy forecasting: integration into electricity grids and markets: a best practice guide," Federal Ministry for Economic Cooperation and Development, Eschborn, 2015. [Online]. Available: [https://energypedia.info/images/2/2a/Discussion\\_Series\\_06\\_Technology\\_web.pdf](https://energypedia.info/images/2/2a/Discussion_Series_06_Technology_web.pdf)
- [10] A. Foley, P. Leahy, A. Marvuglia, and E. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1–8, 2012, doi: 10.1016/j.renene.2011.05.033.
- [11] F. Bouhafs, M. Mackay, and M. Merabti, "Links to the future: communication requirements and challenges in the smart grid," in *IEEE Power and Energy Magazine*, vol. 10, no. 1, pp. 24–32, Jan.-Feb. 2012, doi: 10.1109/MPE.2011.943134.
- [12] H. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012, doi: 10.1016/j.rser.2012.02.049.
- [13] M. Şahin, "Comparison of modelling ANN and ELM to estimate solar radiation over Turkey using NOAA satellite data," *International Journal of Remote Sensing*, vol. 34, no. 21, pp. 7508–7533, 2013, doi: 10.1080/01431161.2013.822597.
- [14] S. Salcedo-Sanz, C. Casanova-Mateo, A. Pastor-Sánchez, D. Gallo-Marazuela, A. Labajo-Salazar, and A. Portilla-Figuera, "Direct solar radiation prediction based on soft-computing algorithms including novel predictive atmospheric variables," *Intelligent Data Engineering and Automated Learning – IDEAL 2013*, pp. 318–325, 2013, doi: 10.1007/978-3-642-41278-3\_39.
- [15] S. Salcedo-Sanz, C. Casanova-Mateo, A. Pastor-Sánchez, and M. Sánchez-Girón, "Daily global solar radiation prediction based on a hybrid coral reefs optimization – extreme learning machine approach," *Solar Energy*, vol. 105, pp. 91–98, 2014, doi: 10.1016/j.solener.2014.04.009.
- [16] M. Lazzaroni, S. Ferrari, V. Piuri, A. Salman, L. Cristaldi, and M. Faifer, "Models for solar radiation prediction based on different measurement sites," *Measurement*, vol. 63, pp. 346–363, 2015, doi: 10.1016/j.measurement.2014.11.037.
- [17] S. Shamshirband, K. Mohammadi, H. Chen, G. Narayana Samy, D. Petković, and C. Ma, "Daily global solar radiation prediction from air temperatures using kernel extreme learning machine: A case study for Iran," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 134, pp. 109–117, 2015, doi: 10.1016/j.jastp.2015.09.014.
- [18] S. Shamshirband, K. Mohammadi, P. Yee, D. Petković, and A. Mostafaeipour, "A comparative evaluation for identifying the suitability of extreme learning machine to predict horizontal global solar radiation," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1031–1042, 2015, doi: 10.1016/j.rser.2015.07.173.
- [19] A. Aybar-Ruiz et al., "A novel grouping genetic algorithm–extreme learning machine approach for global solar radiation prediction from numerical weather models inputs," *Solar Energy*, vol. 132, pp. 129–142, 2016, doi: 10.1016/j.solener.2016.03.015.
- [20] Y. Wu, and J. Wang, "A novel hybrid model based on artificial neural networks for solar radiation prediction," *Renewable Energy*, vol. 89, pp. 268–284, 2016, doi: 10.1016/j.renene.2015.11.070.
- [21] Y. Feng, N. Cui, Q. Zhang, L. Zhao, and D. Gong, "Comparison of artificial intelligence and empirical models for estimation of daily diffuse solar radiation in North China Plain," *International Journal of Hydrogen Energy*, vol. 42, no. 21, pp. 14418–14428, 2017, doi: 10.1016/j.ijhydene.2017.04.084.
- [22] B. Dong, C. Cao, and S. Lee, "Applying support vector machines to predict building energy consumption in tropical region," *Energy and Buildings*, vol. 37, no. 5, pp. 545–553, 2005, doi: 10.1016/j.enbuild.2004.09.009.
- [23] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, "Applying support vector machine to predict hourly cooling load in the building," *Applied Energy*, vol. 86, no. 10, pp. 2249–2256, 2009, doi: 10.1016/j.apenergy.2008.11.035.
- [24] V. Vapnik, *The nature of statistical learning theory*, Springer New York, NY, 2000, doi: 10.1007/978-1-4757-3264-1.
- [25] J. Shi, W. -J. Lee, Y. Liu, Y. Yang, and P. Wang, "forecasting power output of photovoltaic systems based on weather classification and support vector machines," in *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064–1069, May-June 2012, doi: 10.1109/TIA.2012.2190816.
- [26] J. H. Yousif, and H. A. Kazem, "Modeling of daily solar energy system prediction using soft computing methods for Oman," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 13, no. 3, pp. 237–244, 2016, doi: 10.19026/rjaset.13.2936.
- [27] H. A. Kazem, and J. Yousif, "Comparison of prediction methods of photovoltaic power system production using a measured dataset," *Energy Conversion and Management*, vol. 148, pp. 1070–1081, 2017, doi: 10.1016/j.enconman.2017.06.058.
- [28] A. Dairi, F. Harrou, Y. Sun, and S. Khadraoui, "Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach," *Applied Sciences*, vol. 10, no. 23, p. 8400, Nov. 2020, doi: 10.3390/app10238400.
- [29] V. De, T. Teo, W. Woo, and T. Logenthiran, "Photovoltaic power forecasting using LSTM on limited dataset," *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2018, doi: 10.1109/isgt-asia.2018.8467934.
- [30] T. Hiyama, and K. Kitabayashi, "Neural network based estimation of maximum power generation from PV module using environmental information," in *IEEE Transactions on Energy Conversion*, vol. 12, no. 3, pp. 241–247, Sept. 1997, doi: 10.1109/60.629709.
- [31] M. Abuella, and B. Chowdhury, "Solar power forecasting using artificial neural networks," *2015 North American Power Symposium (NAPS)*, 2015, pp. 1–5, doi: 10.1109/NAPS.2015.7335176.
- [32] S. Sulaiman, R. Rahman, I. Musirin, and S. Shaari, "An artificial immune-based hybrid multi-layer feedforward neural network for predicting grid-connected photovoltaic system output," *Energy Procedia*, vol. 14, pp. 260–264, 2012, doi: 10.1016/j.egypro.2011.12.927.
- [33] L. Breiman, "Heuristics of instability and stabilization in model selection," *The Annals of Statistics*, vol. 24, no. 6, 1996, doi: 10.1214/aos/1032181158.
- [34] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*, pp. 1-15, 2000, doi: 10.1007/3-540-45014-9\_1.





- [35] Y. Grandvalet, "Bagging equalizes influence," *Machine Learning*, vol. 55, no. 3, pp. 251–270, 2004, doi: 10.1023/b:mach.0000027783.34431.42.
- [36] T. Chen, and C. Guestrin, "XGBoost", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, doi: 10.1145/2939672.2939785.

## BIOGRAPHIES OF AUTHORS







**Debani Prasad Mishra**     received the B.Tech. in electrical engineering from the Biju Patnaik University of Technology, Odisha, India, in 2006 and the M.Tech in power systems from IIT, Delhi, India in 2010. He has been awarded a Ph.D. degree in power systems from Veer Surendra Sai University of Technology, Odisha, India, in 2019. He is currently serving as Assistant Professor in the Dept of Electrical Engg, IIIT Bhubaneswar, Odisha. His research interests include soft Computing techniques application in power systems, signal processing, and power quality. He can be contacted at email: debani@iiit-bh.ac.in.







**Subhrajit Jena**     is pursuing Bachelor of Technology (B.Tech) in the stream of Electrical and Electronics Engineering in International Institute of Information Technology Bhubaneswar (IIIT-BH), Odisha, India. His research interests include renewable energy, Power grid, machine learning and Artificial Intelligence . He can be contacted by email: subhrajitjena4488@gmail.com.







**Rudranarayan Senapati**     received the B.Tech. in electrical engineering from the Utkal University, Odisha, India, in 2001 and the M.Tech in Communication System Engineering in 2008 from KIIT University, Odisha. He has been awarded a Ph.D. degree in Electrical Engineering from KIIT deemed to be University, Patia Bhubaneswar, Odisha, India, in 2018. He is currently serving as Assistant Professor in the School of Electrical Engg, KIIT, deemed to be University, Patia, Bhubaneswar, Odisha. His research interests include Solar forecasting, Blockchain technology, Renewable Integration to power systems and power quality. He can be contacted at email: rsenapatifel@kiit.ac.in.



**Atman Panigrahi**     is pursuing Bachelor of Technology (B.Tech) in the stream of Electrical and Electronics Engineering in International Institute of Information Technology Bhubaneswar (IIIT-BH), Odisha, India. His research interests include renewable energy, power electronics and power grid. He can be contacted at email: b319009@iiit-bh.ac.in, or atman.panigrahi2001@gmail.com.



**Surender Reddy Salkuti**     received the Ph.D. degree in electrical engineering from the Indian Institute of Technology, New Delhi, India, in 2013. He was a Postdoctoral Researcher with Howard University, Washington, DC, USA, from 2013 to 2014. He is currently an Associate Professor with the Department of Railroad and Electrical Engineering, Woosong University, Daejeon, South Korea. His current research interests include market clearing, including renewable energy sources, demand response, smart grid development with integration of wind and solar photovoltaic energy sources. He can be contacted at email: surender@wsu.ac.kr.