



Use Of XGBoost Method For Very Short-Term Radiation Forecasting on Adaptive Solar Cells

Hikmat Oka Kusuma¹*

¹ Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Surabaya,
Indonesia

*Correspondence: hikmatkusuma@unesa.ac.id

ARTICLE INFO

Article History:

Submitted/Received 22 May 2025

First Revised 30 June 2025

Accepted 30 June 2025

First Available Online 30 June 2025

Publication Date 30 June 2025

Keyword:

Radiation forecasting, XGBoost, Adaptive solar cells, Renewable energy.

ABSTRACT

Solar radiation constitutes a pivotal determinant influencing the efficacy of solar panels, wherein fluctuations may engender uncertainty regarding the generated power and subsequently affect the stability and dependability of solar power systems. The objective of this research is to formulate a method for the short-term prediction of solar radiation employing the XGBoost algorithm. This methodology encompasses data pre-processing, the implementation of the XGBoost model, and the evaluation of the model utilizing the RMSE, MSE, MAPE, and MAE metrics. The experimental findings reveal that the predictive model exhibits commendable performance, with MAE values recorded at 3.3823, MSE at 4001.60, RMSE at 63.26, and MAPE at 5.09%. Despite the presence of discrepancies between the predicted outcomes and the actual data, the overarching trend in the data demonstrates a commendable level of accuracy, with a mere deviation of 5%. These results suggest that the integration of XGBoost methodologies has the potential to enhance the precision of solar radiation predictions for adaptive solar panel systems. Subsequent research endeavors are anticipated to cultivate more robust models through the utilization of an expanded dataset and alternative machine learning combinatorial models to further refine predictions amidst diverse meteorological conditions.

1. INTRODUCTION

Solar energy constitutes one of the most promising and feasible renewable energy sources for the sustainable generation of electricity [1]. The technology for solar power generation, particularly concerning photovoltaic panels, is undergoing rapid advancements in alignment with the escalating global demand for clean energy solutions [2]. Notwithstanding the inherent advantages of solar energy, such as its ubiquitous availability and negligible environmental impact, the efficiency of photovoltaic systems remains a significant challenge [3]. A principal determinant affecting this efficiency is the quantity of solar energy incident upon solar cells [4]. Fluctuations in atmospheric conditions, including but not limited to clouds, precipitation, fog, and atmospheric pollution, contribute to this variability [5]. These variations introduce uncertainties in the power output produced by solar cells, thereby influencing the stability and reliability of systems designed for solar-powered electricity generation [6]. Consequently, it is imperative to conduct research aimed at predicting and alleviating the effects of solar radiation fluctuations on the performance of solar cells, as this is vital for enhancing efficiency and facilitating the integration of solar energy into contemporary energy frameworks.

Solar panels possess the capability to pursue the solar trajectory throughout the diurnal cycle, facilitated by apparatus known as solar tracking systems, which guarantees that the panels are perpetually aligned to optimize solar irradiance [7]. The efficacy of this technological advancement is profoundly contingent upon the precision of contemporaneous estimations regarding the intensity of solar radiation [8]. Accelerated adjustments in the positioning of panels are rendered feasible through precise forecasts of these fluctuations in radiation, which can ultimately augment energy production. Nonetheless, short-term solar radiation is characterized by considerable dynamism and is subject to a plethora of meteorological variables, including cloud cover, humidity levels, and air quality, thereby complicating the prediction process [9]. The intricate and swift transitions inherent in these patterns frequently elude the capabilities of conventional linear regression forecasting methodologies [10]. Consequently, novel approaches are requisite to enhance the fidelity of solar radiation predictions over exceptionally brief temporal intervals.

Solar photovoltaic systems possess the capability to follow the solar trajectory throughout the diurnal cycle, facilitated by devices known as solar tracking systems, which guarantee that the panels maintain optimal orientation to maximize solar irradiance [11]. The effectiveness of this technological advancement is significantly contingent upon the precision of real-time forecasts regarding the intensity of solar radiation [12]. The expediency of panel orientation adjustments is enhanced by accurate anticipations of fluctuations in solar irradiance, which may ultimately augment energy production [13]. Nonetheless, short-term solar radiation exhibits a high degree of variability and is subject to a multitude of meteorological influences, including cloud cover, humidity levels, and air pollutants, rendering accurate prediction a formidable challenge [14]. The intricate and swift alterations in these phenomena are frequently too complex to be effectively captured by conventional linear regression forecasting methodologies [15]. Consequently, innovative approaches are requisite to enhance the precision of solar radiation predictions over exceedingly short temporal intervals [16].

Hammad Ali Khan et al. elucidated the transformative potential of quantum machine learning in the domain of renewable energy forecasting by surmounting the constraints inherent in conventional methodologies, particularly in contexts characterized by limited data availability. In their investigation, the application of deep learning techniques for solar radiation forecasting was employed. Murugesan S., M. Mahasree, et al. innovated a solar energy prediction framework predicated on machine learning, incorporating post-processing optimization and data integrity, and attained an absolute error percentage of 4.7% and 6.3%

under cold and hot conditions, respectively [17]. In this manuscript, Rajnish et al. ascertained that the hyperparameter-optimized deep learning GA-CNN algorithm exhibited superior performance relative to alternative methodologies such as LSTM, KNN-SVM, and CNN-RNN in the forecasting of photovoltaic energy, as determined by performance metrics including RMSE, MAE, MSE, and R-Square [18].

Huang et al. formulated an optimal random forest model for predisolar cells, attaining an accuracy increase of 13.48% over support vector regression while elucidating donor-acceptor configurations and essential constituents such as benzene-1,2-diamine and sulfur-nitrogen bonds, which are pivotal for enhancing operational efficiency [19]. uses the extreme gradient gain (XGBoost) regression algorithm to predict solar power, utilising historical hourly solar radiation data from the city of Johannesburg as the training dataset. Obiora C. N. Ali A et al. used the extreme gradient gain (XGBoost) regression algorithm to predict solar power, utilising historical hourly solar radiation data from the city of Johannesburg as the training dataset [20].

The principal contribution of this scholarly investigation lies in the forecasting of short-term solar radiation utilizing the XGBoost algorithm [21]. The XGBoost methodology has rendered substantial advancements in the forecasting of very short-term solar radiation (15–60 minutes in advance), particularly within the framework of solar energy systems [22]. Its proficiency in managing non-linear and intricate datasets renders it more advantageous than linear or single-tree models, especially in enhancing the precision of real-time predictions.

With respect to its adaptability, XGBoost demonstrates the proficiency to handle diverse forms of input including meteorological data, temporal variables, and historical radiation metrics [23]. Furthermore, this algorithm offers a feature importance assessment, which facilitates the identification of the most significant variables, thus enhancing the comprehension of localized atmospheric dynamics.

In summary, the contributions of XGBoost encompass enhanced precision, computational efficacy, versatility in application, and interpretability, all of which hold significant relevance in the advancement of contemporary intelligent and adaptive solar energy systems[24].

2. METHODS

In the present investigation, the Extreme Gradient Boosting (XGBoost) technique was employed as the primary algorithm for forecasting ultra-short-term solar radiation values, encompassing a predictive time frame of 15 to 60 minutes in advance. The rationale for the selection of XGBoost was predicated on its proficiency in modeling non-linear and intricate interrelationships among input variables, in addition to its demonstrated capacity to attain elevated accuracy levels in a variety of renewable energy forecasting research endeavors.

XGBoost represents a sophisticated ensemble algorithm known as gradient boosting decision tree (GBDT), which operates by incrementally constructing predictive models. Each subsequent model that is developed aims to rectify the prediction inaccuracies generated by its predecessor. This iterative process continues until a level of optimal performance is attained. Furthermore, XGBoost possesses the capability to modify regularization parameters (encompassing both L1 and L2), thereby yielding models that exhibit enhanced stability and a diminished likelihood of overfitting.

In the execution phase, the XGBoost algorithm was employed to analyze input data comprising meteorological variables including temperature, humidity, atmospheric pressure, temporal indicators (hour, day, month), and lagged values from antecedent radiation measurements. Initially, all data underwent normalization utilizing a Standard Scaler to ensure uniformity in the scale across the various features. Subsequently, the dataset is separated into

training and testing portions in an 80:20 ratio, while the XGBoost model is trained with standard parameters and assessed based on predictive metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R^2), and Mean Absolute Percentage Error (MAPE) [25].

This study's utilization of XGBoost is further substantiated by feature importance analysis to ascertain the input variables that most significantly influence the prediction outcomes. This aids in model interpretation and enhances the optimization of predictive systems in practical applications, such as energy management in photovoltaic systems.

XGBoost is regarded as a better method for predicting dynamic, short-term solar radiation because to its high accuracy, computing efficiency, and interpretative capabilities, which are influenced by numerous environmental conditions.

2.1. Material

This research employed a diverse array of materials and ancillary devices, encompassing historical datasets, hardware components, software applications, and programming libraries, to facilitate the development and evaluation of a brief solar radiation forecasting model employing the Extreme Gradient Boosting (XGBoost) methodology. The primary dataset utilized in this research comprises historical global solar radiation measurements acquired from localized sensors or meteorological stations, featuring a temporal resolution of 15 minutes. In conjunction with radiation data, ancillary meteorological variables were also incorporated, including:

- Current data
- Voltage data
- Power
- Radiation
- Air temperature ($^{\circ}\text{C}$)
- Relative humidity (%)
- Wind speed (m/s)
- Time data (hours, days, months)
- Lag value from previous radiation

Figure 1 show the process if collecting solar radiation data. Data was collected over a three-day period from 31 September 2024 to 2 October 2024, on the 4th floor of Building A8, Faculty of Engineering, Unesa. A total of 76 samples were examined.



Figure 1. The process of collecting solar radiation data during a specific period of time..

2.2 EXTREME GRADIENT BOOSTING (XGBOOST)

XGBoost is a machine learning approach based on ensemble learning that integrates many decision trees sequentially through gradient boosting methods. Figure 2 show the flowchart of the XGBoost method that applied in this reasearch. This method involves constructing each subsequent tree to enhance the residual error of its predecessor, hence increasing the model's accuracy with each iteration. The advantage of XGBoost lies in:

- Its ability to handle non-linear and complex relationships between features,
- High computational efficiency due to training processes carried out with gradient-based optimisation,
- Support for regularisation (L1 and L2) to avoid overfitting.

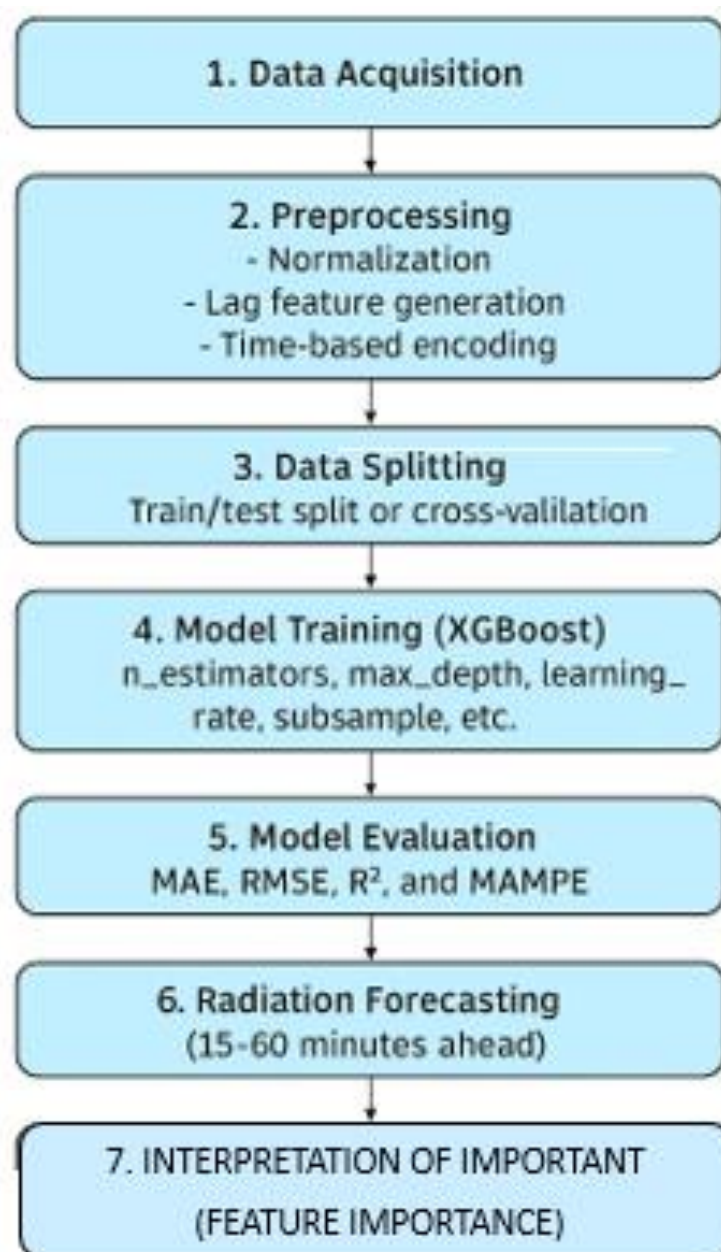


Figure 2. An overview of the method XGBOOST in this experiment.

2.3 Evaluation

This study employed XGBoost to forecast short-term solar radiation values utilizing meteorological and temporal information. The model was trained on historical data with an 80% training ratio and a 20% testing ratio. The performance evaluation utilized MAE, RMSE, R^2 , and MAPE indicators [26].

1. MSE

$$MSE = \frac{1}{M} \sum_{i=1}^m (X_i - Y_i)^2 \quad (1)$$

MSE calculates the average square difference between actual values and predicted values. This metric is sensitive to outliers because it uses the square of the error.

2. MAE

$$MAE = \left(\frac{1}{M}\right) \sum_{i=1}^m |X_i - Y_i| \quad (2)$$

MAE is the average of the absolute differences between the actual values and the predicted values. This metric provides a more direct interpretation of the error compared to MSE. RMSE

$$RMSE = \sqrt{\left(\frac{1}{M}\right) \sum_{i=1}^m (X_i - Y_i)^2} \quad (3)$$

3. RMSE is the square root of MSE and provides an interpretation on the same scale as the original data, making it easier to understand in practical terms. MAPE

$$MAPE = 100\% \times \left(\frac{1}{M}\right) \sum_{i=1}^m \left| \frac{Y_i - X_i}{Y_i} \right| \quad (4)$$

MAPE measures relative error as a percentage, making it useful for comparing model performance on data sets with different scales.

3. RESULTS AND DISCUSSION

3.1. Results

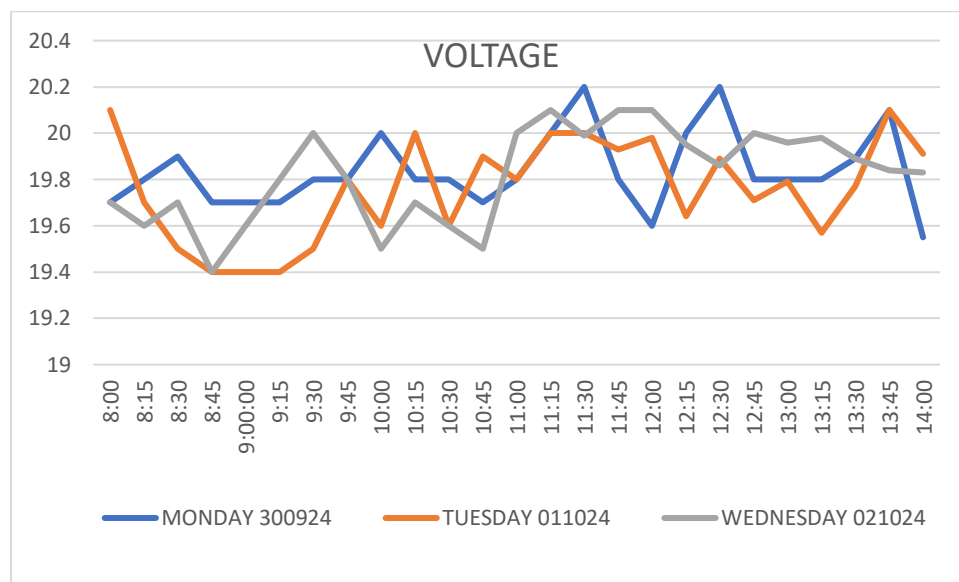


Figure 3. Voltage Data Graph

Figure 3. Voltage Data Graph. This figure illustrates the voltage measurements recorded during the observation period. It shows the electrical potential generated by the solar panel system, which fluctuates based on the level of solar radiation.

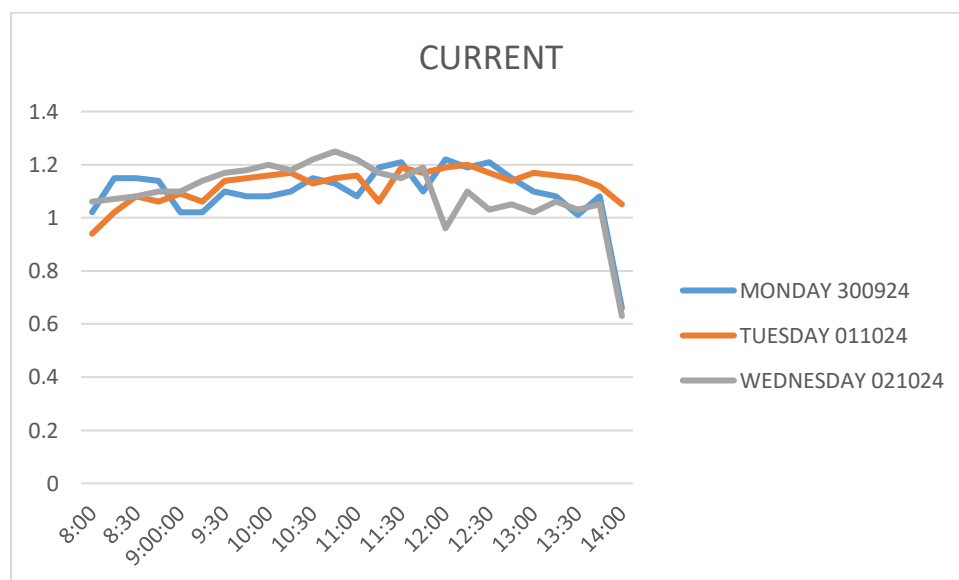


Figure 4. Current FLOW Data Graph

Figure 4. Current Flow Data Graph. This graph displays the electric current (amperage) flowing through the system. Variations in current values are directly affected by changes in solar irradiance and environmental conditions.

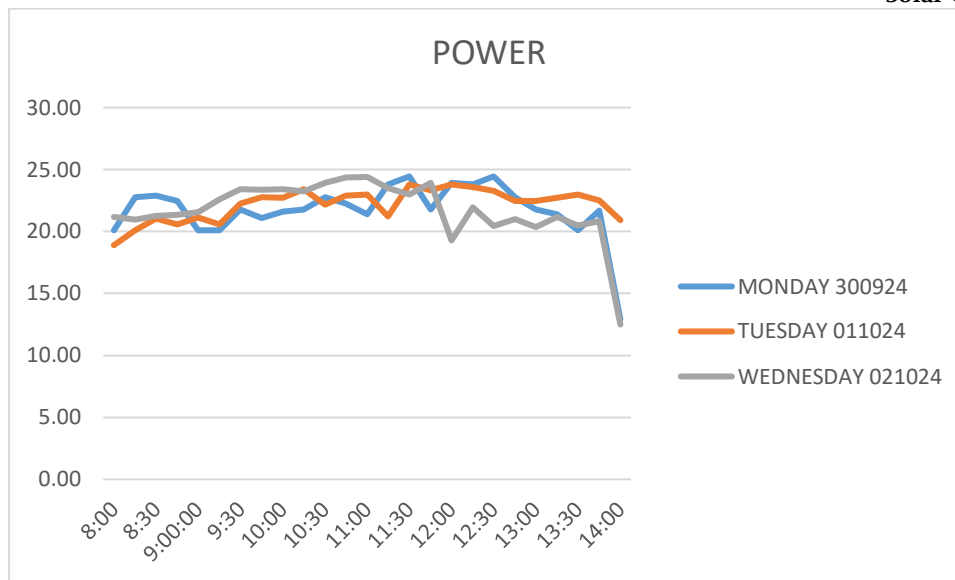


Figure5. Power Graph current

Figure 5. Power Graph. This chart presents the electrical power output of the solar system. Power is the product of voltage and current, and its consistency indicates stable system performance under the observed conditions.

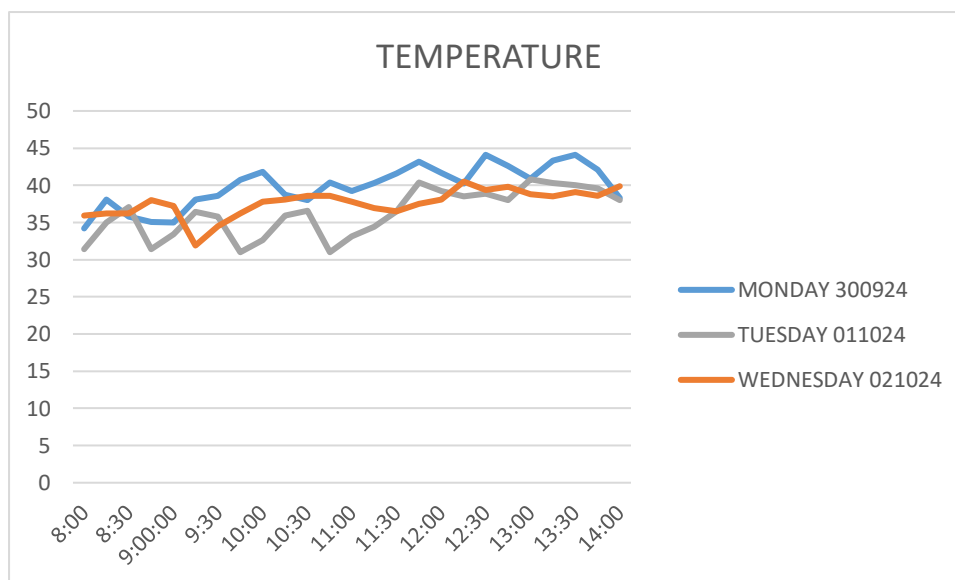


Figure 6. Temperature Graph

Figure 6. Temperature Graph. This graph shows the ambient temperature during the data collection period. Temperature affects the efficiency of photovoltaic cells and is a critical factor in solar radiation modeling.

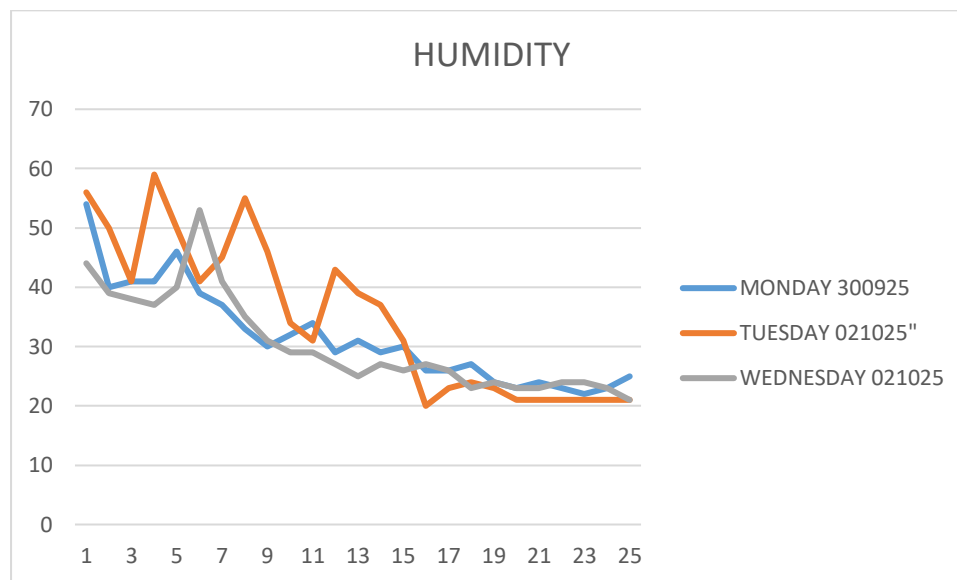


Figure 7. Humidity Graph

Figure 7. Humidity Graph. This figure represents the relative humidity levels. High humidity can impact solar irradiance due to increased water vapor, influencing both power output and prediction accuracy.

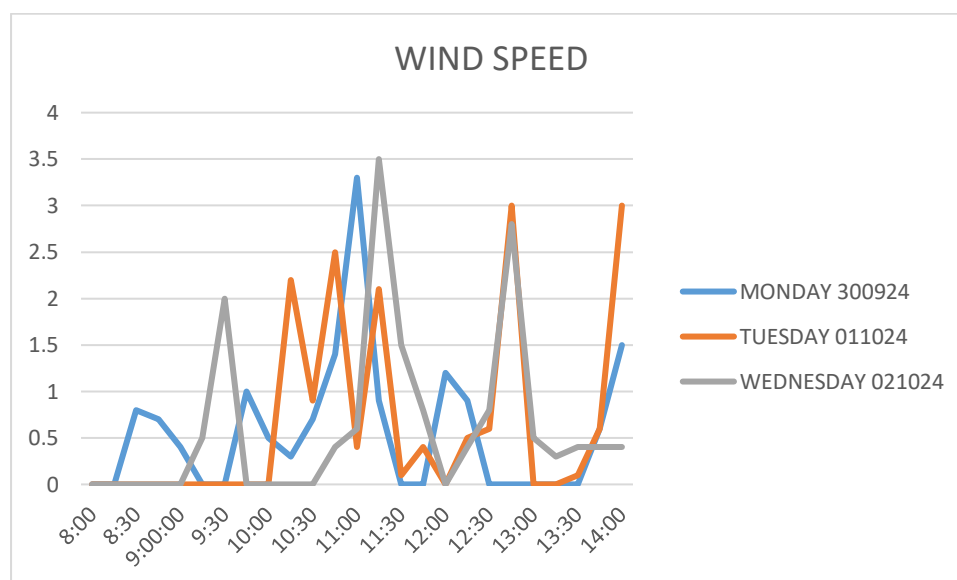


Figure 8. Velocity Wind Graph

Figure 8. Wind Speed Graph. This chart depicts wind speed variations. Wind can affect panel temperature regulation and dust dispersion, indirectly influencing solar panel performance.

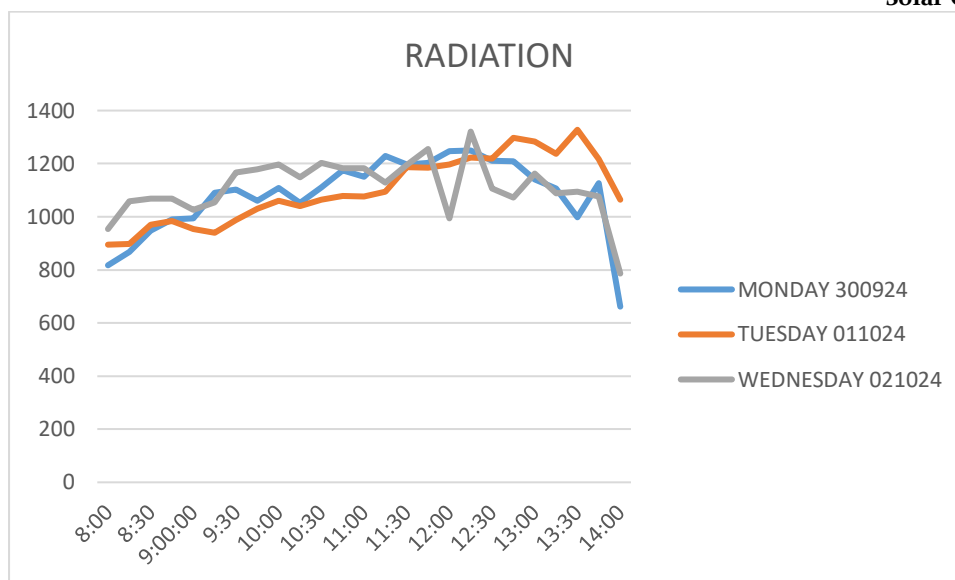


Figure 9. Irradiance Graph

Figure 9. Irradiance Graph. This graph presents the solar radiation (irradiance) measurements recorded by sensors. It reflects the actual sunlight intensity received and serves as the primary variable for prediction.

This data was recorded over three consecutive days, namely on 29, 30 September, and 1 October 2024. This dataset provides a comprehensive overview of the environmental and operational conditions during this period and will serve as the basis for training the XGBoost model. .

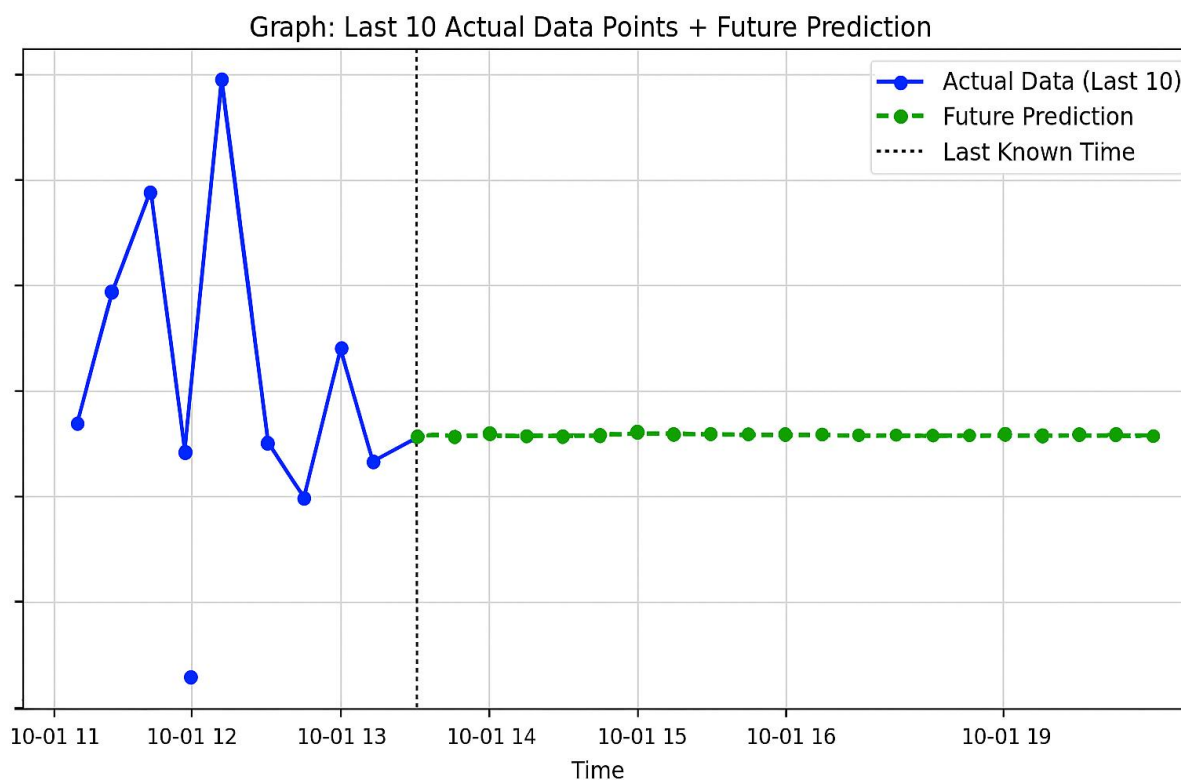


Figure 10. Graph prediction

Figure 10. Prediction Graph. This figure compares actual and predicted solar radiation values using the XGBoost model. The prediction closely follows the actual data trend, especially between 13:45 and 15:00, indicating high model accuracy. According to the attached graph, our radiation forecasts remain accurate until a maximum of 15:00. The graph post-15:00 is not suitable for prediction due to its repetitive nature beyond that time. Table 1 shows the projected solar radiation forecast from 13:45 to 15:00 (in 15-minute intervals).

Table 1. Projected Solar Radiation Forecast from 13:45 to 15:00

Time Predicted	Radiation ($\pm$)
13:45	1110 W/m ²
14:00	1111 W/m ²
14:15	1111 W/m ²
14:30	1112 W/m ²
14:45	1111 W/m ²
15:00	1110 W/m ²

The XGBoost model predicts stable values in the range of 1109–1112 W/m² without extreme fluctuations, indicating consistent clear/normal conditions and making it very useful for regulating solar power system output or load in the near future.

4. CONCLUSION

The investigation and application of a predictive model employing the Extreme Gradient Boosting (XGBoost) algorithm reveal its remarkable effectiveness in anticipating very short-term solar radiation levels, particularly within a 15 to 60-minute interval. XGBoost proficiently discerns complex non-linear correlations between meteorological variables and time with exceptional accuracy, demonstrating reliable performance marked by low prediction error values as per model evaluation results in table 2.

Table 2. Model Evaluation Results

MODEL	Result
MSE	4001.60
RMSE	63.26
R²	0.69
MAPE	5.09%

The main advantages of XGBoost in this context include computational efficiency, versatility with various input data formats, and the ability to evaluate results through feature importance analysis. This model demonstrates reliability in real-time prediction scenarios, as proven by consistent forecast results between 13:45 and 15:00, featuring stable and realistic radiation values.

These findings corroborate earlier research, including that of Rajnish et al. [18], which indicated that deep learning models, specifically GA-CNN, exhibited enhanced performance in photovoltaic energy forecasting relative to conventional approaches. Moreover, Obiora et al. [20] discovered that XGBoost had superior accuracy in predicting solar irradiance utilizing historical data, hence affirming its proficiency in addressing short-term variability across various meteorological situations. Likewise, the research of Cortez et al. [22] demonstrates that XGBoost surpassed ARIMA and LSTM in extremely short-term solar forecasting, particularly in contexts where computational efficiency and real-time responsiveness are paramount.

XGBoost is thus advocated as an efficient and cost-effective predictive technique for data-driven solar energy systems, especially in the realms of power management, dynamic load regulation, and energy storage system oversight during brief time intervals. This research substantially aids the adoption of artificial intelligence in the renewable energy sector, while concurrently fostering opportunities for further development in IoT-based systems and edge computing.

5. AUTHORS' NOTE

I hereby declare that the research, analysis, and findings presented in this study are the result of my own independent work. All sources and references used have been properly cited in accordance with academic standards. I confirm that this research is free from plagiarism and has not been copied or reproduced from any other work without proper acknowledgment.

6. ACKNOWLEDGEMENT

The author expresses sincere gratitude to the Department of Electrical Engineering, Universitas Negeri Surabaya, for providing research facilities and support. Appreciation is also extended to the data acquisition team for their valuable assistance during the experimentation phase.

7. AUTHOR'S CONTRIBUTION

Hikmat Oka Kusuma : Conceptualization, Methodology, Investigation, Writing Original Draft, Supervision, Review and Editing, and Data curation.

7. REFERENCES

- [1] E. T. Sayed *et al.*, "Renewable Energy and Energy Storage Systems," 2023. doi: 10.3390/en16031415.
- [2] K. Abdulmawjood, S. Alsadi, S. S. Refaat, and W. G. Morsi, "Characteristic Study of Solar Photovoltaic Array under Different Partial Shading Conditions," *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3142168.
- [3] S. Iqbal, S. N. Khan, M. Sajid, J. Khan, Y. Ayaz, and A. Waqas, "Impact and performance efficiency analysis of grid-tied solar photovoltaic system based on installation site environmental factors," *Energy Environ.*, vol. 34, no. 7, 2023, doi: 10.1177/0958305X221106618.
- [4] R. J. Mustafa, M. R. Gomaa, M. Al-Dhaifallah, and H. Rezk, "Environmental impacts on the performance of solar photovoltaic systems," *Sustain.*, vol. 12, no. 2, 2020, doi: 10.3390/su12020608.
- [5] J. Yin, A. Molini, and A. Porporato, "Impacts of solar intermittency on future photovoltaic reliability," *Nat. Commun.*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-18602-6.
- [6] O. A. Al-Shahri *et al.*, "Solar photovoltaic energy optimization methods, challenges and issues: A comprehensive review," 2021. doi: 10.1016/j.jclepro.2020.125465.
- [7] S. Pranav, S. Kumar S, S. Biju, L. Monachan, J. Joy, and B. Bobby, "An Integrated System for Monitoring & Control of Solar Panel using IoT & Machine Learning," in *Proceedings of the 2nd International Conference on Modern Trends in Engineering Technology and Management*, 2023. doi: 10.21467/proceedings.160.53.
- [8] A. Mohammad and F. Mahjabeen, "Revolutionizing Solar Energy: The Impact of Artificial Intelligence on Photovoltaic Systems," *Int. J. Multidiscip. Sci. Arts*, vol. 2, no. 3, 2023, doi: 10.47709/ijmdsa.v2i1.2599.

- [9] A. Aslam, N. Ahmed, S. A. Qureshi, M. Assadi, and N. Ahmed, "Advances in Solar PV Systems; A Comprehensive Review of PV Performance, Influencing Factors, and Mitigation Techniques," 2022. doi: 10.3390/en15207595.
- [10] D. K. Murugan *et al.*, "Machine learning approaches for real-time forecasting of solar still distillate output," *Environ. Challenges*, vol. 13, 2023, doi: 10.1016/j.envc.2023.100779.
- [11] J. Karthika, D. Louis Fernandes, K. Logesh Krishna, A. Meenakshi Sundaram, and S. Niruj, "Hybrid Dual Axis Solar Tracker with 3 level DC to DC Boost Converter with GSS based Power Tracking," in *8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022*, 2022. doi: 10.1109/ICACCS54159.2022.9785252.
- [12] N. Krishnan, K. R. Kumar, and C. S. Inda, "How solar radiation forecasting impacts the utilization of solar energy: A critical review," 2023. doi: 10.1016/j.jclepro.2023.135860.
- [13] Neeraj, P. Gupta, and A. Tomar, "Multi-model approach applied to meteorological data for solar radiation forecasting using data-driven approaches," *Optik (Stuttg.)*, vol. 286, 2023, doi: 10.1016/j.ijleo.2023.170957.
- [14] F. Naseem *et al.*, "An integrated approach to air pollution modeling from climate change perspective using ARIMA forecasting," *J. Appl. Agric. Biotechnol.*, vol. 2, no. 2, 2017.
- [15] Q. Li, M. Bessafi, and P. Li, "Mapping Prediction of Surface Solar Radiation with Linear Regression Models: Case Study over Reunion Island," *Atmosphere (Basel)*, vol. 14, no. 9, 2023, doi: 10.3390/atmos14091331.
- [16] K. Yan, H. Shen, L. Wang, H. Zhou, M. Xu, and Y. Mo, "Short-term solar irradiance forecasting based on a hybrid deep learning methodology," *Inf.*, vol. 11, no. 1, 2020, doi: 10.3390/info11010032.
- [17] S. Murugesan, M. Mahasree, F. Kavim, and N. Bharathiraja, "Solar Energy Forecasting With Performance Optimization Using Machine Learning Techniques," *Electr. Power Components Syst.*, 2024, doi: 10.1080/15325008.2024.2316245.
- [18] Rajnish, S. Saroha, and M. K. Saini, "Numerical Modeling and Performance Assessment of Machine Learning-Based Solar Photovoltaic Energy Forecasting System," in *Lecture Notes in Electrical Engineering*, 2024. doi: 10.1007/978-981-99-7383-5_29.
- [19] D. Huang *et al.*, "A machine learning prediction model for quantitative analyzing the influence of non-radiative voltage loss on non-fullerene organic solar cells," *Chem. Eng. J.*, vol. 475, 2023, doi: 10.1016/j.cej.2023.145958.
- [20] C. N. Obiora, A. Ali, and A. N. Hasan, "Implementing extreme gradient boosting (xgboost) algorithm in predicting solar irradiance," in *2021 IEEE PES/IAS PowerAfrica, PowerAfrica 2021*, 2021. doi: 10.1109/PowerAfrica52236.2021.9543159.
- [21] O. Bamisile, C. J. Ejayi, E. Osei-Mensah, I. A. Chikwendu, J. Li, and Q. Huang, "Long-Term Prediction of Solar Radiation Using XGboost, LSTM, and Machine Learning Algorithms," in *2022 4th Asia Energy and Electrical Engineering Symposium, AEEES 2022*, 2022. doi: 10.1109/AEEES54426.2022.9759719.
- [22] J. C. Cortez, L. Zenichi Terada, B. V. Barros Bandeira, J. Soares, Z. Vale, and M. J. Rider, "Comparative Analysis of ARIMA, LSTM, and XGBoost for Very Short-Term Photovoltaic Forecasting," in *2023 15th Seminar on Power Electronics and Control, SEPOC 2023*, 2023. doi: 10.1109/SEPOC58810.2023.10322588.
- [23] R. A. Abbasi, N. Javaid, M. N. J. Ghuman, Z. A. Khan, S. Ur Rehman, and Amanullah, "Short Term Load Forecasting Using XGBoost," in *Advances in Intelligent Systems and Computing*, 2019. doi: 10.1007/978-3-030-15035-8_108.

- [24] J. Y. Oh, D. H. Ham, Y. G. Lee, and G. Kim, "Short-term Load Forecasting Using XGBoost and the Analysis of Hyperparameters," *Trans. Korean Inst. Electr. Eng.*, vol. 68, no. 9, 2019, doi: 10.5370/KIEE.2019.68.9.1073.
- [25] O. M. Mbah, C. I. Madueke, R. Umunakwe, and M. N. Agba, "Extreme Gradient Boosting: A Machine Learning Technique for Daily Global Solar Radiation Forecasting on Tilted Surfaces," *J. Eng. Sci.*, vol. 9, no. 2, 2022, doi: 10.21272/jes.2022.9(2).e1.
- [26] F. Mao, Y. Jiang, B. Cao, and D. Wu, "PV prediction model based on XGBoost with ISMA optimization," 2023. doi: 10.1117/12.2684554.