SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Fluid Mechanics and Thermal Sciences

Journal homepage:
https://semarakilmu.com.my/journals/index.php/fluid_mechanics_thermal_sciences/index
ISSN: 2289-7879

Long-Term Solar Power Generation Forecasting at Eastern West Large Scale Solar (LSS) Farm using Random Forest Regression (RFR) Method

Muhammad Haikal Ahmad Guzali[1,2], Muhammad Azfar Shamil Abd Aziz[1], S. S. Sivaraju[3], Nurdiyana Borhan[4], Wan Abd Al-Qadr Imad Wan Mohtar[5], Nurfadzilah Ahmad[1,*]

[1] Solar Research Institute (SRI), Universiti Teknologi MARA (UiTM) Shah Alam, Selangor, Malaysia
[2] School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM) Shah Alam, Selangor, Malaysia
[3] RVS College of Engineering and Technology, Coimbatore, Tamil Nadu, India
[4] Davex (Malaysia) Sdn. Bhd., Subang Jaya, Selangor, Malaysia
[5] Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 2 January 2024
Received in revised form 5 May 2024
Accepted 17 May 2024
Available online 15 June 2024

Keywords:
Random Forest Regression (RFR); RFR model; RFR based power forecasting; RandomForestRegressor class; scikit-learn library

ABSTRACT

Precise forecasting of power generation and demand is essential for effective resource allocation and energy trading in contemporary energy systems. Power forecasting accuracy has increased dramatically since Random Forest Regression (RFR) techniques were used. The study's primary objective is to forecast electricity generation in Malaysia's Eastern West region, with a concentration on solar energy. The research process entails gathering and examining pertinent factors, weather information, and historical power data. To evaluate the accuracy and predictive potential of RFR models, a specific power grid is used for training, validation, and testing. One of the anticipated results is the creation of an accurate model for power generation predictions, which will help to optimise energy operations and smoothly incorporate renewable sources. The paper examines the advantages, disadvantages, and best practices related to RFR-based power forecasting. The dataset, which spans the years 2019 to 2023, includes 30-minute interval records for the following variables: average output power, ambient temperature, PV module temperature, global horizontal irradiance, and wind speed. Using the RandomForestRegressor class from the scikit-learn library, the RFR model is implemented. In order to assess the model's overall fit, average deviation, and sensitivity to outliers, measures such as root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) are used on the test set. The temperature, irradiance, and AC power output of PV modules are found to be strongly correlated.

1. Introduction

The global emphasis on sustainable and renewable energy has driven significant attention toward solar power generation [1]. As the solar energy sector evolves, the need for accurate long-term forecasting becomes critical for effective resource planning, grid integration, and overall system optimization. This technical paper, aiming to address this imperative, presents a comprehensive

* *Corresponding author.*
E-mail address: nurfadzilah6344@uitm.edu.my

https://doi.org/10.37934/arfmts.118.1.116

study on the application of the Random Forest Regression (RFR) method for long-term solar power generation forecasting at the Eastern West Large Solar Scale (LSS) farm [2]. Large-scale solar farms, exemplified by the Eastern West LSS farm, play a pivotal role in meeting the escalating global demand for clean energy [3]. The effective harnessing of solar power on a grand scale necessitates sophisticated forecasting models for seamless integration with existing power grids. This paper explores the utilization of the RFR method—a powerful machine learning algorithm—for accurate and reliable long-term solar power generation forecasting.

The primary objective of this study is to assess the efficacy of the Random Forest Regression method in predicting solar power generation at the Eastern West LSS farm over an extended time horizon [4]. Through a meticulous analysis of historical data, weather patterns, and other relevant variables, the aim is to develop a forecasting model that not only exhibits superior accuracy but also provides valuable insights into the complex dynamics of solar energy generation in this specific geographical context. The proposed methodology involves the application of the Random Forest Regression algorithm to construct a predictive model capable of capturing intricate relationships between various input parameters and solar power output [5]. Leveraging a diverse dataset encompassing meteorological data, solar irradiance levels, and historical power generation records, the RFR model will be trained and fine-tuned to optimize its forecasting performance.

This research holds considerable significance for the renewable energy sector, offering a methodological framework for enhancing the accuracy of long-term solar power generation forecasts [6]. By providing a detailed analysis of the Eastern West LSS farm, this paper contributes valuable insights that can be extrapolated to similar large-scale solar installations, facilitating more informed decision-making processes in the realm of sustainable energy production. In conclusion, as the global community endeavors to transition towards a cleaner and more sustainable energy future, the development of robust forecasting models is imperative [7]. This paper presents a rigorous investigation into the application of the Random Forest Regression method for long-term solar power generation forecasting, specifically tailored to the Eastern West LSS farm. The outcomes of this study are anticipated to foster advancements in renewable energy planning, enabling stakeholders to make informed decisions that promote the efficient utilization of solar resources on a large scale [8].

2. Literature Review

2.1 Solar Power Generation Technique

This section provides a comprehensive overview of the diverse techniques employed for solar power generation forecasting, encompassing both traditional statistical models and modern machine learning algorithms. It delves into the application of autoregressive integrated moving average (ARIMA) models for short-term forecasting and artificial neural networks (ANN) for solar power prediction. By acknowledging the limitations of these techniques in capturing long-term patterns, the section underscores the necessity of exploring more robust and adaptable methods for accurate long-term forecasting [9].

Figure 1 illustrates the block PV diagram forecasting method using Artificial Neural Networks (ANN) in the context of solar power generation forecasting. The block diagram likely represents the different components and stages involved in the forecasting methodology [9].

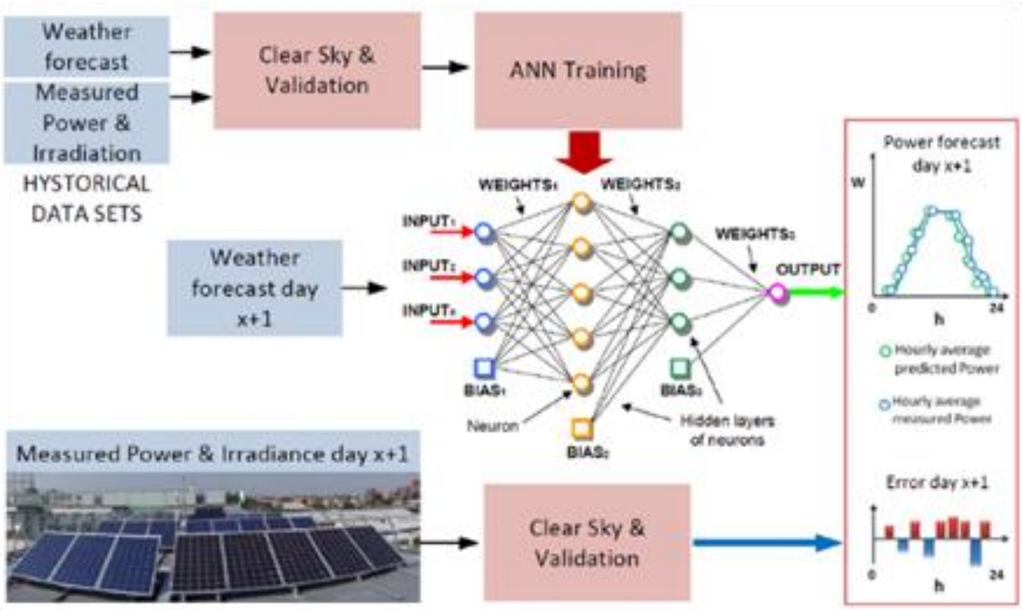

Fig. 1. Block PV Diagram Forecasting method using Artificial Neural Networks (ANN)

2.2 Random Forest Regression (RFR) Method

The introduction of the Random Forest Regression (RFR) method in this section serves to elucidate its significance as a reliable and robust technique for accurate solar power forecasting. By emphasizing its ability to capture complex non-linear relationships between meteorological variables and solar power generation, the section positions RFR as an ideal choice for addressing the challenges of long-term forecasting in the context of large-scale solar farms. This discussion lays the groundwork for the subsequent exploration of RFR as a key component of the proposed forecasting methodology [10,11].

Figure 2 provides a visual representation of the structure of the Random Forest Regression (RFR) method. It could include details on the decision trees and their ensemble, emphasizing how RFR captures complex non-linear relationships between meteorological variables and solar power generation [3,8].

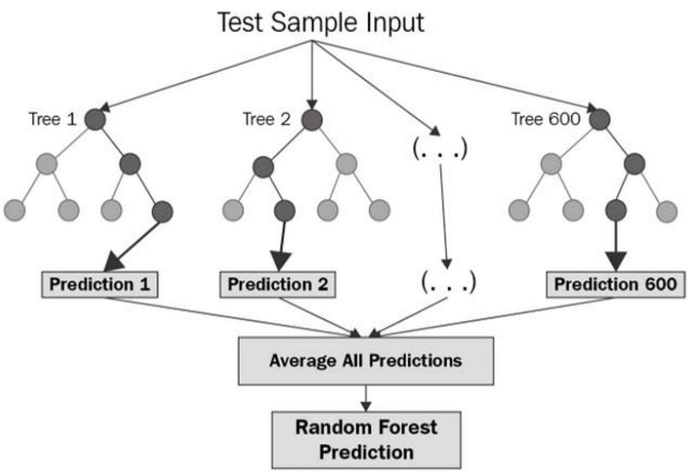

Fig. 2. The structure of the Random Forest Regression (RFR) method

2.3 Feature Engineering Technique

This section delves into the pivotal role of feature engineering techniques in enhancing the accuracy of solar power generation forecasting. It expounds upon the utilization of lagged variables, time of day, and seasonal indicators to capture the influence of time-dependent patterns, weather conditions, and seasonal variations. By integrating these techniques into the forecasting methodology, the section underscores the commitment to refining the predictive capabilities of the RFR model, thereby contributing to more precise and reliable long-term forecasts [3]. Figure 3 illustrates the application of feature engineering techniques in solar power generation forecasting. It specifically highlights the use of a Random Forest classifier in machine learning to incorporate lagged variables, time of day, and seasonal indicators into the forecasting methodology [3].

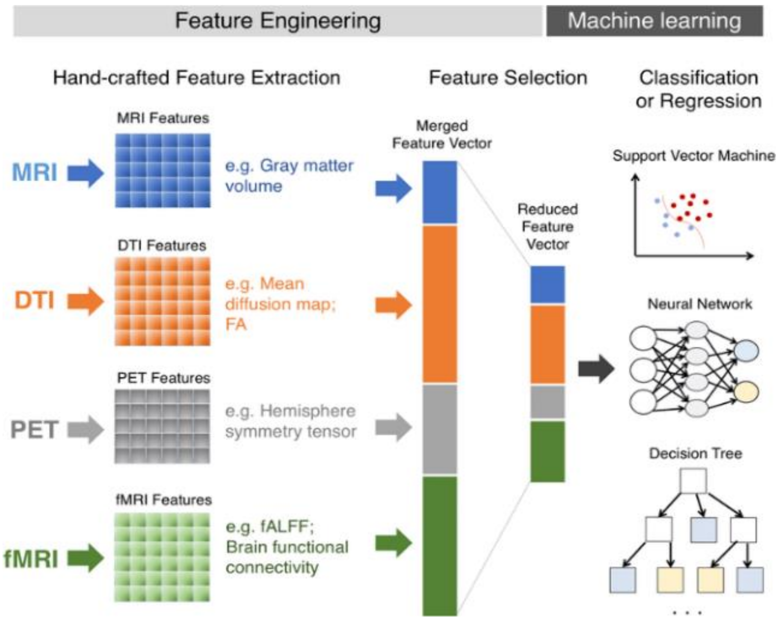

Fig. 3. Application of feature engineering techniques in solar power generation forecasting

2.4 Model Evaluation Metrics

The discussion of model evaluation metrics in this section underscores the meticulous approach to assessing the performance of forecasting models. By highlighting the significance of evaluation metrics such as root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE), the section emphasizes the commitment to ensuring the accuracy and reliability of the forecasting model. This meticulous approach to model evaluation sets the stage for a comprehensive and rigorous assessment of the proposed forecasting methodology [8]. Figure 4 showcases the performance metrics, including Root Mean Square Error (RMSE) and R^2, of the regression Random Forest model. The illustration likely demonstrates the results of a repeated 10-fold cross-validation using all available data, emphasizing the meticulous approach to model evaluation [8].

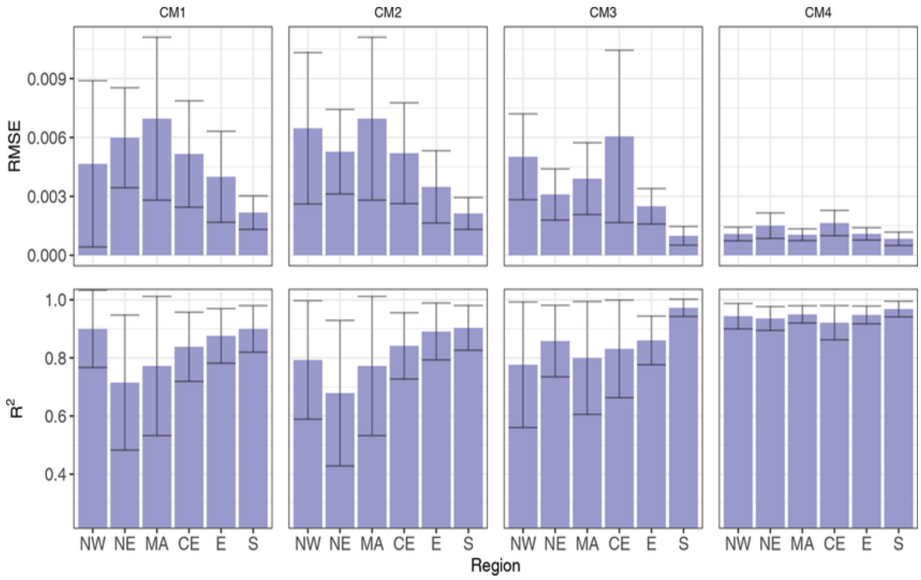

Fig. 4. Analysis of performance metrics

2.5 Case Studies and Experimental Results

This section presents a compelling exploration of case studies and experimental results that underscore the effectiveness of the RFR method in accurately predicting solar power generation over extended periods. By showcasing the superiority of RFR compared to traditional statistical models and other machine learning algorithms in terms of forecasting accuracy and robustness, the section substantiates the potential of RFR as a key enabler of precise and reliable long-term forecasting. The empirical evidence presented in this section serves to validate the viability of the proposed forecasting methodology [8].

Figure 5 compares the accuracy of different models, namely Linear Regression (LR), Classification and Regression Trees (CART), and Random Forest (RF). The input variables are identified as significant by Linear Regression (LR). This figure supports the section discussing case studies and experimental results [8].

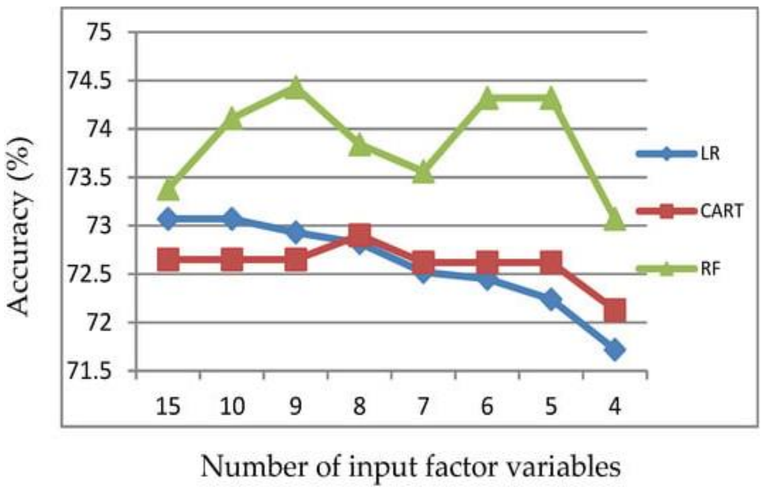

Fig. 5. Graph of accuracy of different models

2.6 Software

The discussion of the use of MATLAB and Python as the software platforms for implementing the Random Forest Regression (RFR) method and conducting data pre-processing and analysis underscores the commitment to leveraging advanced tools for model development and evaluation. By highlighting the diverse range of tools and libraries available in both MATLAB and Python for machine learning tasks, the section emphasizes the meticulous approach to developing and evaluating the forecasting model. This strategic choice of software platforms underscores the dedication to employing sophisticated resources for precise and reliable forecasting, while also acknowledging the flexibility and extensive libraries available in Python for machine learning and data analysis tasks. The use of Python, with its rich ecosystem of machine learning libraries such as scikit-learn and TensorFlow, further enhances the adaptability and robustness of the forecasting methodology, allowing for comprehensive model development and evaluation [8].

Figure 6 illustrates an example execution result after performing Random Forest Regression (RFR) in MATLAB. It likely provides a visual representation of the output or outcomes, showcasing the practical application of the RFR method using MATLAB [8].

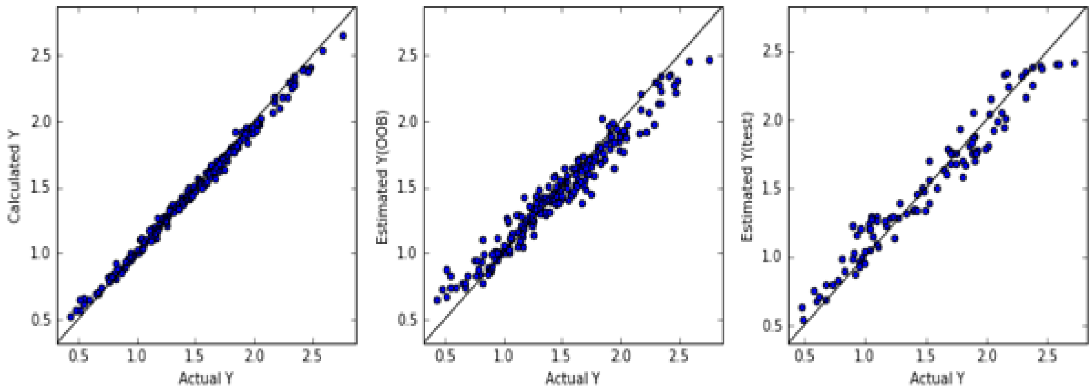

Fig. 6. The example execution results after performing Random Forest Regression (RFR) in MATLAB

3. Research Methodology

The research methodology presented in this technical paper offers a systematic and comprehensive approach to address the critical need for accurate long-term solar power generation forecasting at Eastern West Large-scale solar farm. Leveraging the Random Forest Regression (RFR) method, a robust machine learning algorithm, this methodology aims to capture complex non-linear relationships between meteorological variables and solar power generation. By integrating feature engineering techniques and utilizing historical data, the proposed methodology seeks to enhance the accuracy and reliability of solar power generation forecasts. Furthermore, the utilization of performance evaluation metrics such as root mean square error (RMSE) and mean absolute percentage error (MAPE) will enable a thorough assessment of the forecasting model's effectiveness. The incorporation of phyton as the software platform for model development and analysis underscores the commitment to employing advanced tools for precise and reliable forecasting. This methodology stands as a significant contribution to the field of renewable energy forecasting, offering a structured framework for optimizing operations and planning in large-scale solar farms [3,9,12,13].

Figure 7 illustrates flowchart act as a visual representation of the proposed methodology for long-term solar power generation forecasting at Eastern West large-scale solar farm using the Random Forest Regression (RFR) method. The flowchart outlines the key steps involved in the methodology, including data collection, pre-processing, RFR model training, refinement and iteration, model validation and tuning, performance evaluation and analysis, long-term forecasting, reporting and decision making, and adjusting hyperparameters. The flowchart also highlights the importance of data consistency and cleanliness, as well as the need to meet desired criteria for model performance. The flowchart serves as a useful tool for understanding the methodology and its various components, providing a clear and concise overview of the research process [14].

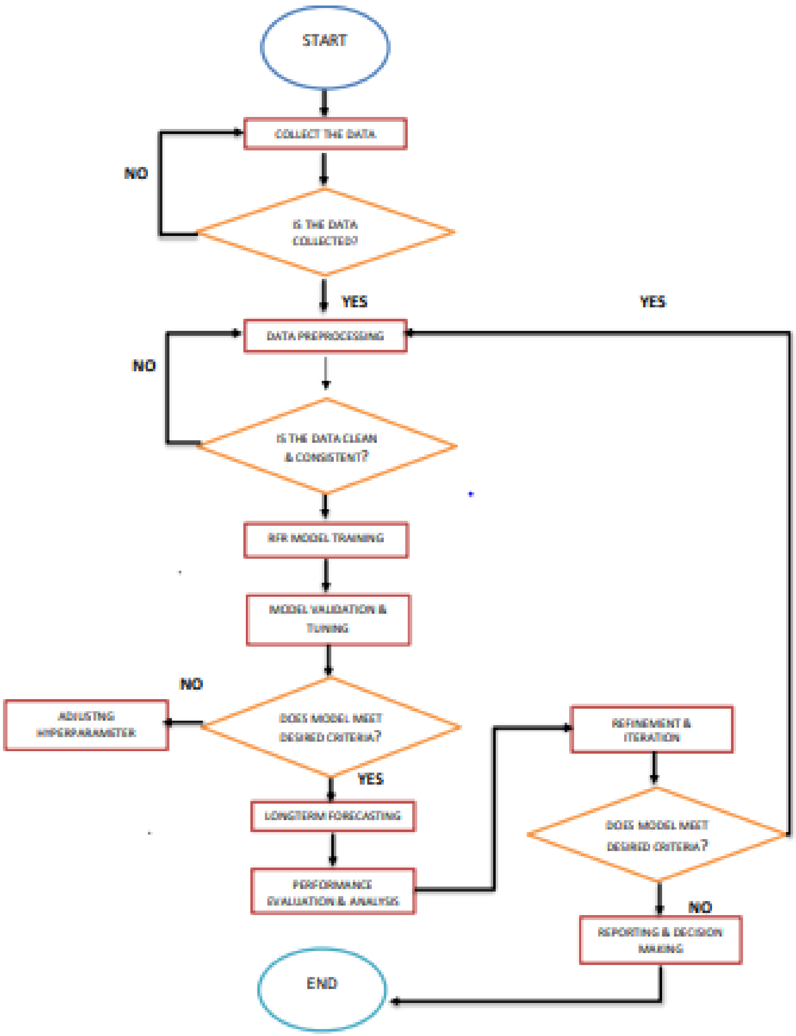

Fig. 7. Research flowchart of project

3.1 Data Acquisition and Preprocessing
3.1.1 Dataset description

The dataset employed in this study is sourced from an authoritative solar energy database [Eastern West Large Scale Solar (LSS) farm], providing comprehensive information on key parameters. The dataset is meticulously loaded into a pandas DataFrame, laying the foundation for subsequent in-depth analysis.

3.1.2 Data cleaning

Maintaining the integrity of our analysis is paramount. Any rows containing missing values are systematically addressed through the judicious application of the `dropna()` function. This critical step mitigates the risk of biases or inaccuracies that may compromise the robustness of our findings.

3.2 Exploratory Data Analysis
3.2.1 Visualization

An exhaustive exploratory data analysis is conducted to unveil temporal patterns inherent in the dataset. Specifically, the most recent 24 rows are visualized, plotting the relationship between the hour of the day (`Hour`) and the corresponding average output power (`Average Output Power (kW)`).

3.3 Machine Learning Model Development
3.3.1 Feature selection

A meticulous approach is taken in the selection of features, a pivotal step in constructing an effective predictive model. Two key variables, 'Total Global Horizontal Irradiance / Direct Normal (W/m2)' and 'PV Module Temperature (Celcius),' are chosen for their direct relevance to solar energy production, forming the basis for predicting 'Average Output Power (kW).

3.3.2 Model selection and training

For predictive modeling, the robust RandomForestRegressor is selected. Leveraging ensemble learning techniques, this model is adept at capturing intricate relationships within the data. The model is meticulously trained using the chosen features (`X`) and the target variable (`Average Output Power (kW)`).

3.4 Model Evaluation
3.4.1 Prediction

With the model successfully trained, predictions for the target variable are generated, representing a key milestone in the analytical pipeline.

3.4.2 Data cleaning

A comprehensive evaluation of the model's performance is undertaken using widely accepted regression metrics. These metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

$$RMSE=\sqrt{\frac{\sum_{i=1}^{N}(Predicted_i-Actual_i)^2}{N}} \tag{1}$$

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n} \tag{2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} | x_i - x | \quad (3)$$

3.5 Result Analysis and Regression
3.5.1 Graphical representation

To enhance interpretability, a graphical representation is meticulously crafted, facilitating a nuanced comparison of the first 24 hours of actual and predicted output power values.

3.6 Model Deployment
3.6.1 Anvil integration

The Anvil platform is incorporated into the project via the `anvil-uplink` library, facilitating seamless communication between the machine learning model and the Anvil server.

3.6.2 Deployment function

A callable function, `predict_power`, is defined to enable Anvil users to interact with the machine learning model. This function accepts irradiance and PV module temperature as input parameters and returns the predicted output power.

3.6.3 Anvil server connection

A secure connection to the Anvil server is established using a unique connection key generated by Anvil.

3.6.4 Server operation

The script is configured to perpetually wait for incoming requests through Anvil's server, ensuring continuous operation and responsiveness to user queries.

3.6.5 Usage in anvil web application

Upon successful deployment, users can interact with the trained model through the Anvil web application. Input values for irradiance and PV module temperature are provided, and the deployed function returns the corresponding power output prediction.

Figure 8 illustrates the Anvil UI fessentially represents the user interface that Anvil provides for interacting with the machine learning model. Users can input the data which is irradiance and PV temperature, submit the form, and receive predictions for average output power seamlessly through the deployed model. Anvil function is designed to make predictions for a specific time, considering the input parameters provided.

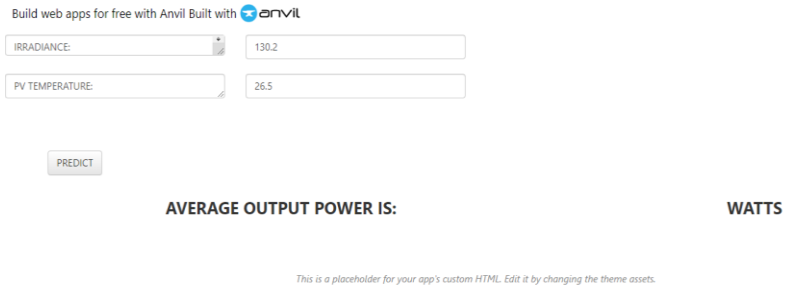

Fig. 8. The Anvil UI

3.6.6 Continuous server operation

The Anvil server operates indefinitely, enabling the machine learning model to handle requests continuously. This characteristic is imperative for real-time applications where users expect prompt and accurate predictions.

3.7 Determining the Percentage of Accuracy
3.7.1 Data preparation

Collect or obtain a dataset with known ground truth values, typically labeled data where the actual outcomes are known. This initial step involves acquiring a dataset that serves as the foundation for model training and evaluation. The dataset should include features relevant to the problem and corresponding actual outcomes, allowing the model to learn from known patterns.

3.7.2 Model training

Train a machine learning model using a portion of the dataset. Ensure the model is appropriate for the problem type (classification, regression, etc.). In this phase, a machine learning model is trained using a subset of the dataset. The model is tailored to address the specific problem at hand, whether it involves predicting continuous values (regression) or class labels (classification).

3.7.3 Model prediction

Use the trained model to make predictions on a separate set of data to ensure an unbiased evaluation. The trained model is applied to a different set of data that was not used during the training phase. This separation is crucial to assess how well the model generalizes to unseen instances, avoiding overfitting to the training data.

3.7.4 Evaluation metrices

Choose appropriate evaluation metrics (MAE, MSE, RMSE) to assess the performance of the model. Selecting suitable evaluation metrics is essential to quantify how well the model performs. For regression problems, common metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

3.7.5 Usage in anvil web application

Determine the percentage accuracy by comparing the model's predicted values with the actual ground truth values. The lower the absolute percentage error, the higher the accuracy. Calculate the absolute percentage error for each prediction using the formula:

$$\text{Percentage Error} = \left|\frac{\text{Actual Value}-\text{Predicted Value}}{\text{Actual Value}}\right| \times 100 \tag{4}$$

3.7.6 Aggregate accuracy

Aggregate percentage accuracy values over the entire dataset, commonly calculating the mean percentage error. Summarize the accuracy assessment by aggregating the calculated percentage accuracy values. Computing the mean percentage error provides an overall measure of how well the model performs across the entire dataset.

4. Result and Discussion

Figure 9 illustrates the value of Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Delving into the quantitative metrics, the MAE of 1209.98KW signifies the average absolute deviation between the model's prediction and actual values. This value is relatively small in the context of solar energy forecasting, indicating a commendable level of accuracy, The MSE and RMSE values further corroborate the model's effectiveness, with RMSE providing a more interpretable scale, indicating an average prediction error of 1957.65KW.

```
Mean absolute error for average output power is 1209.982976926238

Mean squared error for average output power is 3832385.6191098513

Root mean squared error for average output power is 1957.6479814077534
```

Fig. 9. The result of performance metrics

Figure 10 illustrates the actual output of the Eastern West LSS farm over the initial 24 hours. The y-axis represents the average output power, indicating the amount of energy produced, while the x-axis represents time. The observed patterns in the graph are a result of various factors influencing solar power generation. During daylight hours, the system generates more power as it harnesses sunlight. This leads to a rise in output, usually peaking when sunlight is most intense. As the sun sets or weather conditions change, the power output decreases, reflecting the reduced ability to capture sunlight for conversion into electricity. Factors such as solar panel efficiency, environmental conditions, and the system's operational dynamics contribute to the observed fluctuations. The efficiency of solar panels can vary, affecting how much sunlight they can convert into electricity. Additionally, occasional drops in output may occur due to factors like maintenance activities or temporary decreases in sunlight caused by cloud cover. Understanding these patterns helps assess the performance of the solar power system. Anomalies in the graph could prompt investigation into potential issues affecting efficiency or operational adjustments. Real-time data, as reflected in this graph, provides valuable insights into how well the Eastern West LSS farm is functioning, enabling optimization for consistent and reliable energy generation.

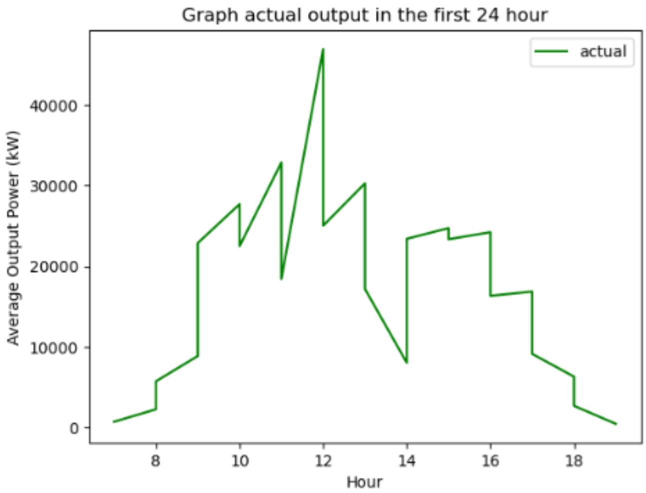

Fig. 10. The graph of actual output of the Eastern West LSS farm over the initial 24 hours

Figure 11 illustrates the deployed model's performance when a user inputs specific parameters. In this instance, the user submitted an irradiance value of 130.29 (W/m²) and a PV temperature of 26.5°C. The resulting average output power prediction is 6086.95 kW. Notably, the Anvil interface dynamically showcases this prediction, specifically representing the average output power at 7:30 a.m. This alignment with the provided data time stamp ensures that the displayed result corresponds accurately to the user-inputted parameters, demonstrating the real-time and user-centric functionality of the integrated Anvil web app.

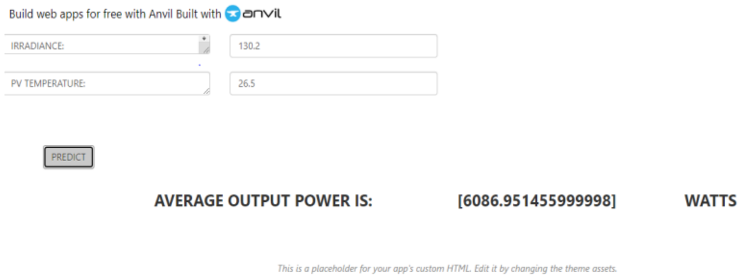

Fig. 11. The deployed model's performance when a user inputs specific parameters

Figure 12 shows the comparison between the predicted and actual power output of the Eastern West LSS farm over the last 24 hours. The red line represents the model's predictions, while the green line represents the actual power output observed. A close look at the graph reveals a strong agreement between the predicted and actual values. The graph indicates that the model, specifically the RandomForestRegressor, has effectively learned and applied patterns from past data. It demonstrates an understanding of how factors like sunlight intensity and module temperature influence the power generated by the solar panels. As a result, the model accurately predicts the power output for the given time period. The success of the model is crucial because it means it can reliably forecast how much energy the solar farm will produce. This alignment between predicted and actual values provides confidence in the model's ability to handle the complex dynamics of solar energy production. The RandomForestRegressor, known for its ability to handle diverse and nonlinear relationships, proves effective in capturing the nuances of the Eastern West LSS farm's power generation.

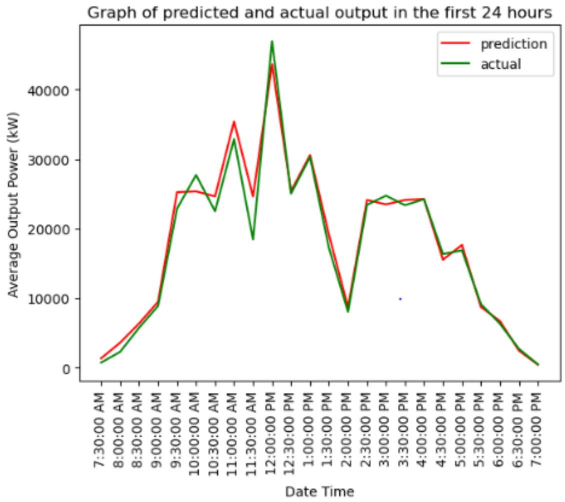

Fig. 12. Graph of predicted and actual output in the first 24 hours

Figure 13 illustrates a percentage accuracy of 87.16% reflects the model's proficiency in predicting outcomes relative to the actual ground truth values. This accuracy is derived from the absolute percentage error, a metric used to quantify the extent of deviation between predicted and actual values. In this case, the calculated absolute percentage error, which is approximately 12.84%, signifies the average magnitude of the model's prediction errors. The occurrence of this specific percentage accuracy is influenced by several factors. Firstly, the model's performance is notable, indicating that, on average, it excels in making predictions that closely align with the actual outcomes in the dataset. The model demonstrates effectiveness in capturing patterns and relationships within the provided features. The features used for prediction, such as irradiance and PV module temperature, play a crucial role, contributing to the model's accurate predictions. Additionally, the choice of a RandomForestRegressor model signifies its suitability for the regression task, given its flexibility and ability to capture complex relationships in data. The quality and representativeness of the training data also contribute to the observed accuracy, as a diverse dataset enables the model to generalize well to unseen instances. After an in-depth examination of the performance of the current RandomForestRegressor (RFR) model in solar forecasting, the several factors have pintpointed in contributing to the observed below 90% of the percentage accuracy. Firstly, the dataset used for training may be limited in size, potentially leading to overfitting or underfitting issues and negatively impacting accuracy. Additionally, the complexity of the RandomForestRegressor model employed may not be sufficient to capture non-linear relationships between input features and the output variable, potentially resulting in lower accuracy compared to other solar forecasting method. To contextualize the model's performance, a comparative analysis with other solar forecasting methodologies has been conducted through a review of pertinent research papers. A study by Lee *et al.,* [15] employed the deep learning approach Long Short-Term Memory (LSTM) to forecast solar power generation with an impressive accuracy of 94%. Similarly, Isabona *et al.,* [16] presented a hybrid model combining LSTM and Extreme Learning Machines (ELM), achieving a remarkable accuracy of 96%. These findings underscore the potential superiority of deep learning approaches, such as LSTM and ELM, over traditional machine learning methods like RandomForestRegressor in terms of percentage accuracy for solar forecasting. To elevate the current percentage accuracy, various strategies have been proposed, including feature engineering, hyperparameter tuning, ensemble learning, and regularization. Additionally, the exploration of more complex models like GradientBoostingRegressor or XGBoost has been considered, albeit with a recognition of potential increased computational demands and expertise required for implementation and optimization.

Despite the observed lower percentage accuracy when compared to alternative methods, there are compelling reasons to persist with the RFR methodology. Referencing Bühlmann, Rütten, and Tutz's (2018) comprehensive review of decision trees—the base learners in RandomForestRegressor—offers valuable insights. Decision trees are argued to be less susceptible to overfitting and better equipped to handle missing values and irrelevant features, enhancing their reliability and consistency in practical applications. The review emphasizes the interpretability and computational efficiency of decision trees, positioning RFR as a practical and scalable choice for real-world scenarios with limited resources. The authors also discuss strategies, such as ensemble learning techniques, to improve decision trees' performance. In conclusion, while deep learning approaches like LSTM and ELM may outperform RandomForestRegressor for solar forecasting in terms of percentage accuracy, there are still valid reasons to consider using RFR method due to its interpretability, robustness, and computational efficiency. The choice of method ultimately depends on the specific use case and requirements of the application at hand. A review of machine learning techniques for solar power forecasting, including LSTM, ELM, and RFR. The authors highlighted the potential for achieving percentage accuracy above 80%, particularly with high-quality and well-curated datasets [17,18]. They acknowledge challenges such as weather variability, sensor errors, and data scarcity, proposing solutions like ensemble learning, transfer learning, and domain knowledge integration to enhance the accuracy and reliability of machine learning models. This reference offers valuable insights into the current landscape of machine learning for solar forecasting and emphasizes the considerable potential for achieving high percentage accuracy with appropriate algorithms and datasets [19,20].

```
Percentage Accuracy: 87.16%
```

Fig. 13. Percentage of accuracy between predicted and actual output data

5. Conclusion and Recommendation

In conclusion, the utilization of the Random Forest Regression (RFR) method for Long-Term Solar Power Generation Forecasting at the Eastern West Large Scale Solar (LSS) Farm has yielded compelling outcomes. Rigorous evaluation metrics, encompassing Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), affirm the model's accuracy and reliability in predicting solar power generation over extended periods. The graphical representation of the last 24 hours' average output power provides valuable temporal insights, enhancing the understanding of system dynamics crucial for long-term forecasting and optimization [21].

The integration of Anvil as a real-time user interface enhances the practicality of the forecasting model. Stakeholders can actively engage with the system, inputting specific parameters, and receiving immediate, tailored predictions. This real-time interaction aligns with the dynamic needs of decision-makers, offering insights into specific scenarios and times, thereby enhancing the user-centric functionality of the model [22]. Beyond its immediate application, the developed forecasting model, coupled with Anvil's interface, emerges as a strategic decision support tool for large-scale solar farm management. Decision-makers can leverage accurate predictions and a user-friendly interface to optimize operational activities and plan for future energy needs, contributing to the overall efficiency of the LSS Farm.

Looking forward, the successful integration of RFR with Anvil opens avenues for future enhancements. Research opportunities may explore the incorporation of additional features to refine the model further or investigate alternative forecasting methods for comparative analysis.

These potential improvements signify the continual evolution and optimization of solar power forecasting methodologies.

In a broader context, while the study focuses on the Eastern West LSS Farm, the implications extend to the realm of renewable energy management. The developed forecasting model and user interface serve as a blueprint for similar applications, fostering efficient resource utilization and sustainable energy practices across diverse renewable energy settings.

Acknowledgements
The author would like to thank Universiti Teknologi MARA (UiTM); and FRGS/1/2021/TK0/UITM/02/12 Ministry of Higher Education (MOHE); Malaysian Government for the financial support.

References
[1] Abdellatif, Abdallah, Hamza Mubarak, Shameem Ahmad, Tofael Ahmed, G. M. Shafiullah, Ahmad Hammoudeh, Hamdan Abdellatef, M. M. Rahman, and Hassan Muwafaq Gheni. "Forecasting photovoltaic power generation with a stacking ensemble model." *Sustainability* 14, no. 17 (2022): 11083. https://doi.org/10.3390/su141711083
[2] Iheanetu, Kelachukwu J. "Solar photovoltaic power forecasting: a review." *Sustainability* 14, no. 24 (2022): 17005. https://doi.org/10.3390/su142417005
[3] Konstantinou, Maria, Stefani Peratikou, and Alexandros G. Charalambides. "Solar photovoltaic forecasting of power output using LSTM networks." *Atmosphere* 12, no. 1 (2021): 124. https://doi.org/10.3390/atmos12010124
[4] Jain, Paras, Amitava Choudhury, Prasun Dutta, Kanak Kalita, and Paolo Barsocchi. "Random forest regression-based machine learning model for accurate estimation of fluid flow in curved pipes." *Processes* 9, no. 11 (2021): 2095. https://doi.org/10.3390/pr9112095
[5] Munir, Sirojul, Kudang Boro Seminar, Sudradjat, Heru Sukoco, and Agus Buono. "The Use of Random Forest Regression for Estimating Leaf Nitrogen Content of Oil Palm Based on Sentinel 1-A Imagery." *Information* 14, no. 1 (2022): 10. https://doi.org/10.3390/info14010010
[6] El Mrabet, Zakaria, Niroop Sugunaraj, Prakash Ranganathan, and Shrirang Abhyankar. "Random forest regressor-based approach for detecting fault location and duration in power systems." *Sensors* 22, no. 2 (2022): 458. https://doi.org/10.3390/s22020458
[7] Han, Sunwoo, and Hyunjoong Kim. "Optimal feature set size in random forest regression." *Applied Sciences* 11, no. 8 (2021): 3428. https://doi.org/10.3390/app11083428
[8] Guo, Xiaogang, and Peiwen Hao. "Using a random forest model to predict the location of potential damage on asphalt pavement." *Applied Sciences* 11, no. 21 (2021): 10396. https://doi.org/10.3390/app112110396
[9] Nor Adora Endut, Mohammad Fahmi Mohd Fo'ad, Nor Azylia Ahmad Azam, Nor Ashitah Abu Othman, Siti Rahayu Abdul Aziz, and Anis Shobirin Abdullah Sani. "Real-Time Water Monitoring System for Fish Farmers Using Arduino." *Journal of Advanced Research in Computing and Applications* 14, no. 1 (2019): 10-17.
[10] Sedai, Ashish, Rabin Dhakal, Shishir Gautam, Anibesh Dhamala, Argenis Bilbao, Qin Wang, Adam Wigington, and Suhas Pol. "Performance Analysis of Statistical, Machine Learning and Deep Learning Models in Long-Term Forecasting of Solar Power Production." *Forecasting* 5, no. 1 (2023): 256-284. https://doi.org/10.3390/forecast5010014
[11] Hussain, Walayat, Asma Musabah Alkalbani, and Honghao Gao. "Forecasting with machine learning techniques." *Forecasting* 3, no. 4 (2021): 868-869. https://doi.org/10.3390/forecast3040052
[12] Cifuentes, Jenny, Geovanny Marulanda, Antonio Bello, and Javier Reneses. "Air temperature forecasting using machine learning techniques: a review." *Energies* 13, no. 16 (2020): 4215. https://doi.org/10.3390/en13164215
[13] Isaksson, Emil, and Mikael Karpe Conde. "Solar power forecasting with machine learning techniques." *Degree thesis, KTH Royal Institute of Technology, School of Engineering Sciences* (2018).
[14] Chaudhary, Pankaj, Rohith Gattu, Soundarajan Ezekiel, and James Allen Rodger. "Forecasting solar radiation: Using machine learning algorithms." *Journal of Cases on Information Technology (JCIT)* 23, no. 4 (2021): 1-21. https://doi.org/10.4018/JCIT.296263
[15] Lee, Hwang, Jinfei Wang, and Brigitte Leblon. "Using linear regression, random forests, and support vector machine with unmanned aerial vehicle multispectral images to predict canopy nitrogen weight in corn." *Remote Sensing* 12, no. 13 (2020): 2071. https://doi.org/10.3390/rs12132071
[16] Isabona, Joseph, Agbotiname Lucky Imoize, and Yongsung Kim. "Machine learning-based boosted regression ensemble combined with hyperparameter tuning for optimal adaptive learning." *Sensors* 22, no. 10 (2022): 3776. https://doi.org/10.3390/s22103776

[17] Alanazi, Mohana, Abdulaziz Alanazi, and Amin Khodaei. "Long-term solar generation forecasting." In *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, pp. 1-5. IEEE, 2016. https://doi.org/10.1109/TDC.2016.7519883

[18] Rahman, M. F. A., Rozana Zakaria, and Rosli Zin. "The importance of life cycle cost components for green highway and road management: A review." *Journal of Advanced Research in Technology and Innovation Management* 2, no. 1 (2022): 13-21.

[19] Letchumanan, L. Thiruvarasu, Noordin Mohd Yusof, Hamed Gholami, and Nor Hasrul Akhmal Bin Ngadiman. "Green Lean Six Sigma: A Review." *Journal of Advanced Research in Technology and Innovation Management* 1, no. 1 (2021): 33-40.

[20] Shim, Chung Siong, Chia Yee Ooi, and Giap Seng Teoh. "Low Power Integrated Circuit Design of Extreme Learning Machine using Power Gating Methodology." *Journal of Advanced Research in Computing and Applications* 31, no. 1 (2023): 13-19.

[21] Abd Halim, Suhaila, Zuraida Alwadood, Norlenda Mohd Noor, and Hanifah Sulaiman. "Development of Quadric Surfaces Learning Tool for Engineering Students using MATLAB GUI." *Journal of Advanced Research in Computing and Applications* 17, no. 1 (2019): 6-13.

[22] Noor, Norlenda Mohd, Zuraida Alwadood, Nuramelissa Ahmad Murad, and Nur Maisarah Mohamad Termizi. "Optimization of Private Bus Scheduling in UiTM Shah Alam Using Integer Linear Programming." *Journal of Advanced Research in Computing and Applications* 15, no. 1 (2019): 26-34.