Physics of semiconductor devices & renewable energies journal

Vol. 02 n. 01 (2025); ISSN 2992-040X
Journal homepage: www.psdrej.net

Accurate Solar Power Forecasting: A Comparative Study of machine Learning Techniques

H. Ammour[1], W. Tercha[2], F. Chekired[3], S. Tadjer[2], R. Rouas[1]

[1]Laboratory of Advanced Technologies in Elzctrical Engineering and Computer Science, Mouloud Mammeri University, Tizi ouzou , Algeria, [2]Electrification of Industrial Enterprises Laboratory, University of Boumerdes, Boumerdes 35000, Algeria. [3] Unité de Développement des Équipements Solaires, UDES, Centre de Développement des Energies Renouvelables, CDER, Tipaza 42004. Tipaza, Algeria.

*Corresponding author *Email: f.chekired@udes.dz*

Abstract

Effective real-time energy management is crucial for the transition to a sustainable and resilient energy system. Solar energy, characterized by its variability due to weather conditions, seasonality, and geographic location, requires reliable forecasting models to maximize its integration and efficient management within the grid. Anticipating energy consumption is also vital to avoid demand peaks and imbalances in electricity distribution. This study deals with machine learning techniques, such as random forest models, decision trees, and XG Boost for modeling and predicting solar energy production and energy consumption by comparing theirs ability to handle complex datasets, identify nonlinear relationships, and provide accurate forecasts.

Keywords: *Energy management, machine learning, forecasting, solar energy*.

1. Introduction

The energy transition is essential for a sustainable future, and it is based on optimizing the production, storage and distribution of energy while integrating renewable sources. This transition is not only about reducing our dependence on fossil fuels, but also about minimizing greenhouse gas emissions to combat climate change. Improving energy efficiency, developing smart infrastructure and promoting the adoption of green technologies are key to achieving these ambitious goals [1].

The application of artificial intelligence (AI) in energy management brings new and promising perspectives. AI techniques, such as machine learning can predict energy needs, proactively manage resources, and optimize the performance of energy systems. AI can also help with early fault detection and predictive maintenance, reducing the risk of breakdowns and improving infrastructure reliability [2].

A. energy management

Energy management encompasses a set of procedures and methods to monitor control and optimize the use of energy resources. It includes the planning, monitoring and regulating energy flows to ensure a maximum efficiency, to meet the needs in a sustainable way, and to minimize environmental impacts [3] [4].
An optimal energy management strategy improves the reliability of the system. These strategies, based on sensor inputs and advanced information technology, provide optimal resource planning. These methodologies are intended to maximize output power and longevity, while decreasing operational and environmental expenses.

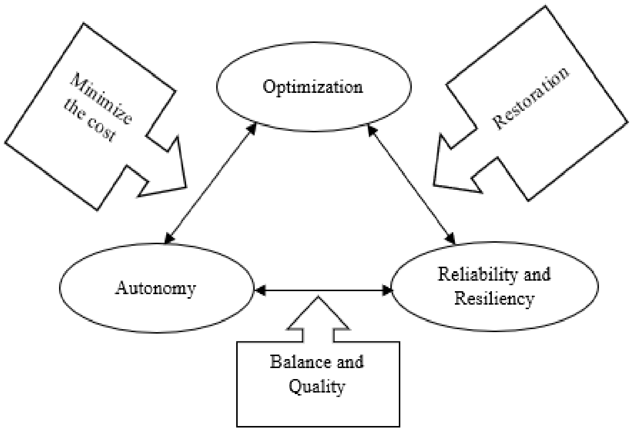

Figure 1: Types of problems solved by energy management

B. Energy management strategies

Energy management strategies in electricity grids are designed to optimize the use and distribution of energy. Some common strategies include:

- Control of the energy demand:

The goal of managing energy demand is to control consumption in order to balance supply and demand and maximize available resources and the effectiveness of the electrical system. This involves shifting payments to low-cost periods, utilizing intelligent technologies, and providing financial incentives to reduce consumption during peak periods, which enables energy savings and (reduce energy loss) [5][6].

- Optimization of renewable energy:

The goal of renewable energy optimization is to maximize the use of internal energy sources like solar and wind power while maintaining dependable and efficient electrical system operation. This depends on forecasting energy output, integrating that output with demand, and utilizing storage technologies to maintain a balance between supply and demand. The forecasts enable the network to adjust its operation to meteorological circumstances, while energy storage allows excess energy to be stored for later use.

- Energy storage:

The management of energy storage maximizes battery utilization to meet electrical system requirements. This includes monitoring charge and discharge timing in accordance with rates, projected renewable energy production, and planned consumption. The management of charge cycles extends the battery's life cycle. The network is stabilized by the batteries by lowering the pics of charge. A clear integration with other system components is necessary for optimal operation [7].

- Load management :

Ensuring the continuous operation of critical equipment, such as medical devices and security systems, is the primary responsibility of loads management. Loads that are time-sensitive and necessary for comfort are given equal priority. This prioritization minimizes service interruptions, maximizes the use of available resources, and ensures reliable and efficient energy provisioning [8].

2. Motivation of the study

The transition to more sustainable and resilient energy systems is a global priority, particularly due to the need to reduce greenhouse gas emissions and meet growing energy demand. Micro-grids, which are localized electrical networks, play a crucial role in this transition by enabling the integration of renewable energy sources, such as solar and wind, while offering greater flexibility and energy independence [9].

However, the effective management of these micro-grids presents major challenges, primarily due to the inherent variability of renewable energy sources. For example, solar energy production is highly dependent on weather conditions and can fluctuate unpredictably. This variability makes it difficult to balance energy supply and demand, which can lead to inefficiencies and instabilities in the grid [10].

It is in this context that energy prediction techniques based on machine learning become essential. These techniques allow for the development of predictive models capable of providing accurate estimates of energy production and consumption. By anticipating these fluctuations, it becomes possible to better manage energy resources, avoid demand peaks, and optimize the use of available resources [11].

The integration of these techniques into micro-grid management can improve grid reliability and efficiency, reduce operational costs, and facilitate the integration of renewable energy. Therefore, this study aims to explore advanced machine learning methods for energy prediction in a micro-grid, in order to contribute to more optimized and sustainable energy management [10].

3. Methodology

Machine learning, a key field of artificial intelligence, has revolutionized computing by offering solutions to complex problems that traditional methods cannot solve. Its strength lies in its ability to detect nonlinear patterns and complex relationships within data. Unlike classical linear models, machine learning algorithms can analyze diverse and unpredictable data, adapting to fluctuations that occur in real-world contexts, such as weather conditions or economic trends. This flexibility enhances the accuracy of predictions, making these techniques particularly useful in fields like energy forecasting, where anticipating variations in solar or wind production is crucial for balancing energy supply and demand. Consequently, the adoption of machine learning algorithms has significantly increased across various sectors, as they enable more informed decision-making through precise and reliable data analysis.

A. Random forest model

The random forest is a supervised ensemble learning algorithm that is particularly effective for energy management in microgrids. It is based on the combination of several base models, typically decision trees, which are merged to accurately predict energy needs. Unlike linear regression, the random forest is better suited to the complexities of energy data, overcoming the limitations of traditional models through the judicious selection of predictive variables. This approach reduces model variability and improves prediction accuracy [12].

The operation of the random forest relies on aggregating the predictions of multiple decision trees. Each tree is trained on a random sample of the training dataset, ensuring diversity among the models. At each node, a random subset of the available features is selected, which contributes to the diversity of the trees and reduces the correlation between them. When a new observation is introduced, it is passed through each individual decision tree, and the final prediction is obtained by aggregating the results from each tree, either by averaging (for regression) or by majority voting (for classification). The performance of the random forest is then evaluated by comparing the predictions obtained with the actual values in a test dataset, and hyperparameter optimization techniques can be applied to further enhance the model's performance [13] .

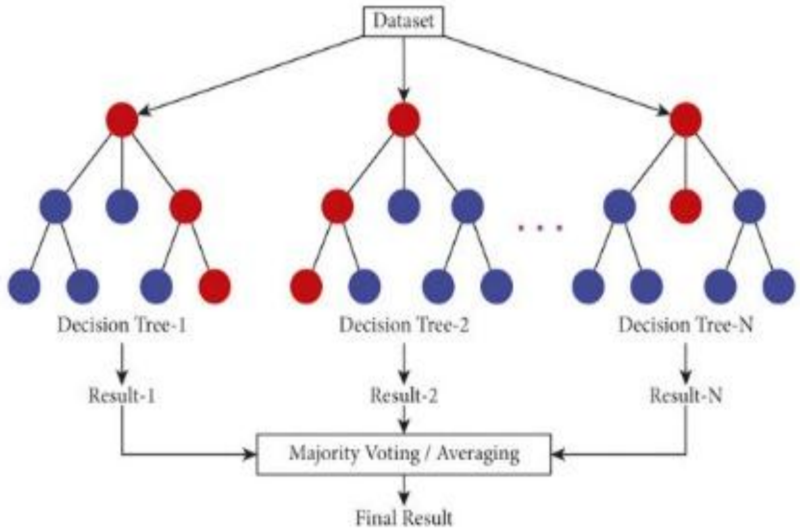

Figure 2: Illustration of the random forest model

B. decision tree model

A decision tree is a predictive model structured in the shape of a tree, where each internal node represents a feature or attribute of the data, each branch represents an outcome of a test on that attribute, and each leaf represents a class label or a prediction value. This model uses a set of rules derived from the data's characteristics to make decisions or predictions [15].

The functioning of decision trees is based on the progressive segmentation of a dataset into smaller subsets, using specific features to guide each division. The algorithm begins by selecting the feature that offers the best separation between the different classes or categories of data, then splits the dataset into subsets based on the values of that feature. This process is repeated recursively for each newly created subset, choosing at each step the best feature to split the remaining data. This division continues until a stopping criterion is reached, such as a maximum tree depth or a minimum number of observations in the leaves.

Once the decision tree is built, it can be used to predict the class or output value for new observations. By following the corresponding path in the tree based on their characteristics, new observations end up at a leaf where the prediction is obtained. This approach allows decision trees to capture complex patterns in the data and provide interpretable predictions for a wide range of classification and regression problems [16] [17].

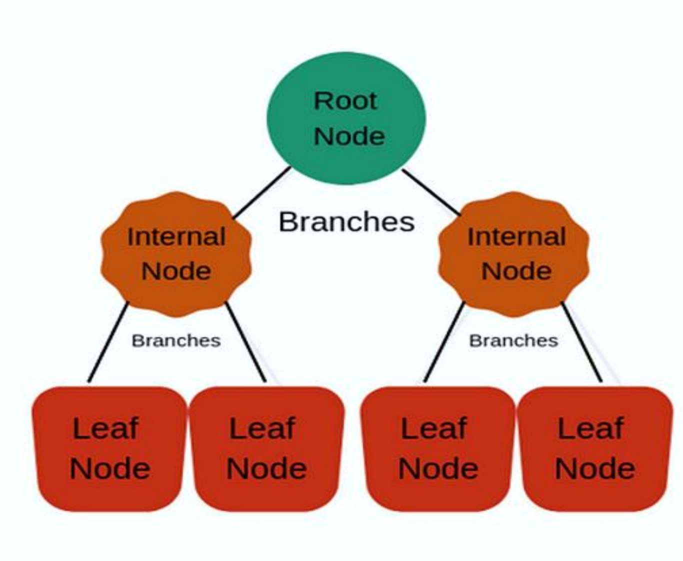

Figure 3: Illustration of decision tree model

C. Gradient boosting model

Extreme Gradient Boosting (XGBoost) is a highly efficient, open-source machine learning library widely used for solving regression and classification problems. Designed to enhance traditional boosting algorithms, XGBoost employs an iterative approach to build predictive models by sequentially adding weak estimators, typically decision trees, and refining them to minimize a specified loss function.

This technique, known as "gradient boosting," enables XGBoost to achieve accurate predictions by optimizing the loss function at each iteration. It incorporates built-in regularization to reduce overfitting, efficient parallelization to handle large datasets, and flexibility to adapt to a range of machine learning tasks. Due to its precision, speed, and versatility, XGBoost is extensively applied across various domains such as finance, healthcare, marketing, and scientific research [18]

XGBoost operates using a method called gradient boosting, which involves iteratively combining multiple weak prediction models to form a more powerful model. Initially, XGBoost constructs a simple model, often a decision tree, to predict the target. By comparing the predictions of this initial model with the actual values, it calculates the residuals, which are the prediction errors. A new model is then built to predict these residuals, adjusting the weights of examples based on the previous model's prediction errors. This new model is combined with the previous one to form a more robust model. This process is repeated iteratively until a stopping criterion is met, such as a fixed number of iterations or a lack of significant performance gains. Once the training is complete, the ensemble of models is used to make predictions on new data. Through this iterative approach and loss function optimization at each step, XGBoost constructs accurate and robust predictive models suitable for a wide range of machine learning tasks [14] [19].

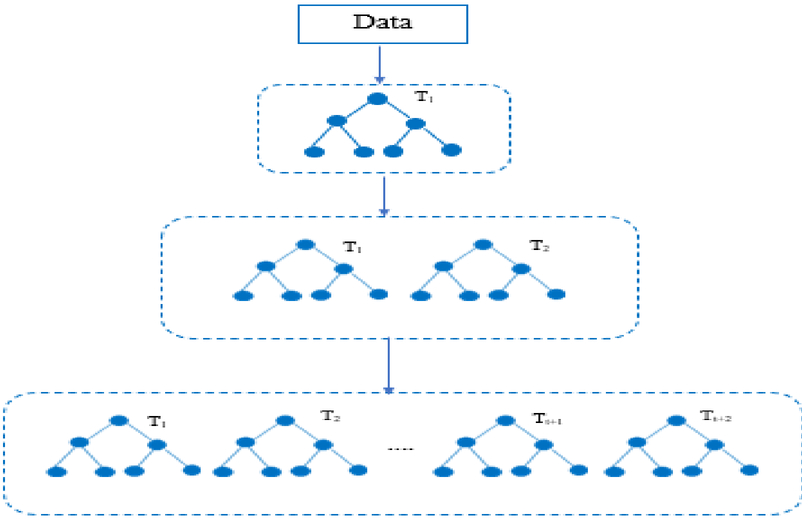

Figure 4: Illustration of gradient boosting model

D. Accuracy indicators

Forecast accuracy plays a fundamental role in decision-making, particularly in sectors such as energy, finance, and the economy. It enables the anticipation of variations, optimization of resources, and minimization of inefficiencies. By reducing the gap between expected and actual outcomes, accuracy improves the management of fluctuations and adjusts strategies to real needs, thereby enhancing system efficiency while promoting more sustainable resource management.

Accuracy indicators are essential tools for evaluating the performance of forecasting models. They quantify the reliability of forecasts by measuring the gaps with observed results, allowing errors to be detected and models to be refined. These measures provide valuable information for adjusting forecasts to data realities, ensuring more informed decision-making and continuous system optimization, which is crucial in critical domains. In the following equations we consider that:

y_i: Is the actual value.

$\hat{y}_i$: Is the predicted value

n: Is the number of observations.

- Mean Absolute Error (MAE)

Mean Absolute Error is a widely used measure to quantify the accuracy of forecasts. It measures the average magnitude of errors in a set of predictions, providing a numerical value that represents the deviation between the forecasts and actual values. MAE is calculated by summing the absolute differences between the predicted values and the actual values and dividing by the number of observations. [20] [21]

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (1)$$

- Mean Squared Error (MSE)

Mean Squared Error measures the dispersion of predictions relative to actual values, indicating lower accuracy if it is high. It is calculated by averaging the squares of the differences between predicted and actual values, making it more sensitive to outliers due to the squaring in the calculation [20] [22]

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

- Root Mean Squared Error **(**RMSE**)**

Root Mean Squared Error is a commonly used metric to evaluate the accuracy of a forecasting or prediction model. It is calculated by taking the square root of the average of the squared differences between actual values and predicted values. This provides a measure of the average deviation between actual and predicted values, expressed in the same units as the original data. In other words, RMSE indicates the dispersion of prediction errors around the regression line. The lower the RMSE value, the more accurate the model [21]

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2} \quad (3)$$

4. Case study

The data used in this case study is related to solar energy production in Spain for the period from the 1st january to the 29th february 2024. This period was selected for its representativeness and relevance in the study of energy production models. The data collected was carefully reviewed to understand the patterns of solar production, seasonal variations and any other factors influencing the availability of solar energy. By analyzing these data, we seek to identify trends, fluctuations and solar energy production cycles in Spain. This preliminary analysis will to better understand the overall behaviour of solar production and lay the foundations for the application of different modelling techniques. Our goal is to draw strong conclusions about the variability of energy production Spain during this period, which will then allow us to apply the the most reliable prediction method, to anticipate solar energy production in other contexts.

	Month	Day	Hour	Minute	Power
0	1	1	0	0	28.0
1	1	1	0	15	28.0
2	1	1	0	30	28.0
3	1	1	0	45	28.0
4	1	1	1	0	28.0

Figure 5: Visualisation of solar power production for all the 1st January

	Month	Day	Hour	Minute	Power
count	5760.000000	5760.000000	5760.000000	5760.000000	5760.000000
mean	1.483333	15.516667	11.500000	22.500000	3155.842361
std	0.499766	8.685015	6.922788	16.771966	4681.208413
min	1.000000	1.000000	0.000000	0.000000	0.000000
25%	1.000000	8.000000	5.750000	11.250000	28.000000
50%	1.000000	15.500000	11.500000	22.500000	148.000000
75%	2.000000	23.000000	17.250000	33.750000	5712.000000
max	2.000000	31.000000	23.000000	45.000000	16448.000000

Figure 6: Visualisation of solar power production for all the 1st January

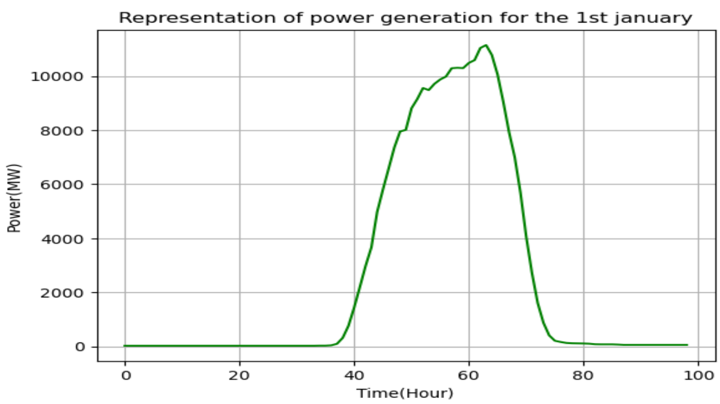

Figure 7: Visualisation of solar power production for all the 1st January

A crucial step was to evaluate the performance of the model. To do this, a comprehensive display of test data was performed, along with the values predicted by the algorithm, which included machine learning techniques such as random forest (RDF), XGBoost (Extreme Gradient Boosting) and decision trees. This approach has allowed us to closely examine the capacity of model to generalize from training data and produce accurate predictions on test data. The comparative analysis between actual values and predictions of the model was essential to assess its reliability and ability to meet the needs of this study. In addition, the metric errors have been calculated to quantitatively evaluate the performance of the model.

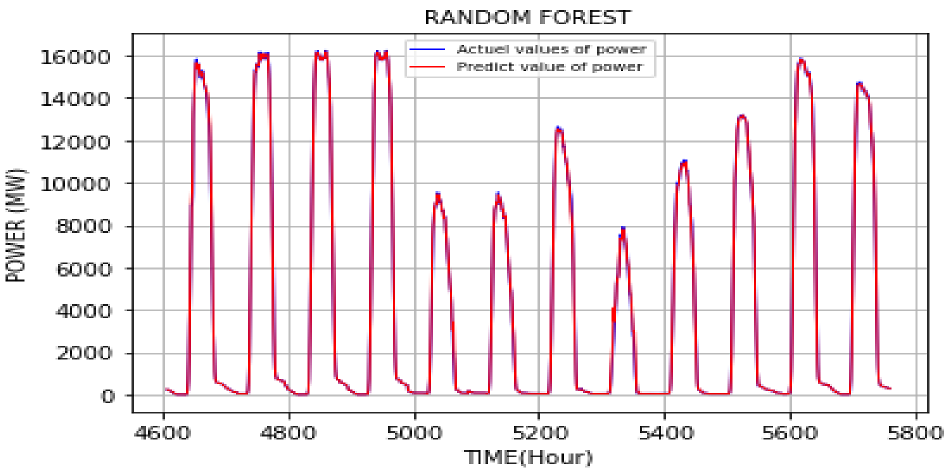

Figure 8: The application of the Random Forest model

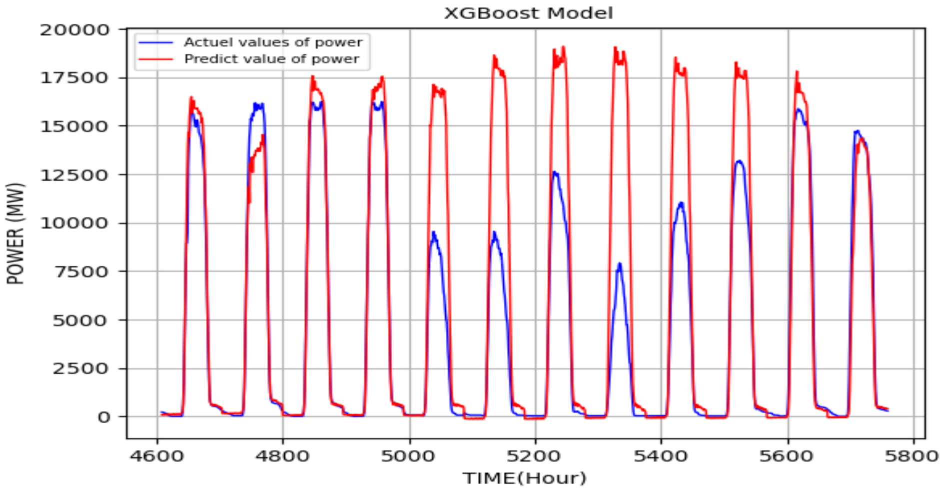

Figure 9: The application of the XGBoost model

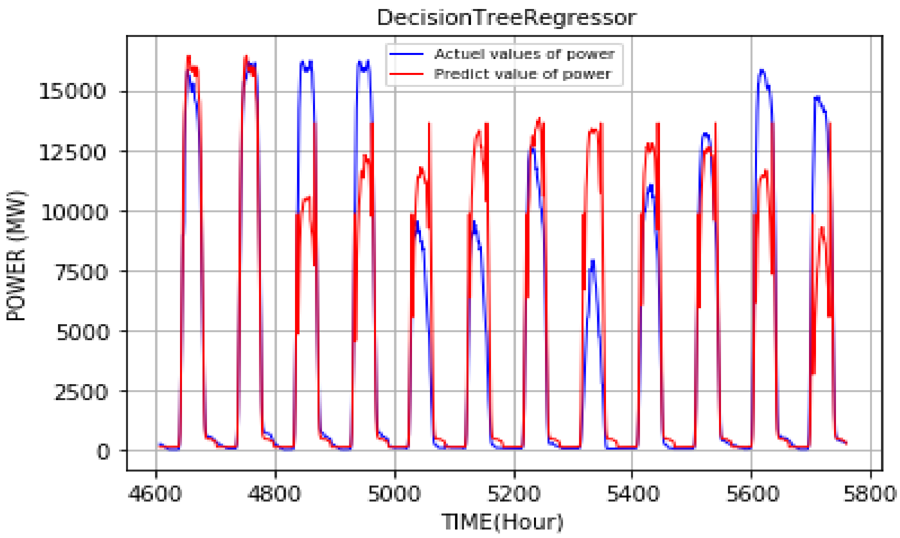

Figure 10: The application of the Decision Tree model

5. Results and iscussion

After running our code, Table (1) shows the various error metrics calculated, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These measures provide precise indications of the performance and accuracy of our prediction models.

MSE evaluates the average of the squared differences between predicted values and actual values. MAE measures the average of the absolute differences between predicted values and actual values. RMSE is a measure of mean squared error that takes into account the square root of MSE, thereby providing an indication of the dispersion of errors.

The analyse of these metrics allows the determination of which model provides the most reliable and accurate performance for our application.

Table 1: Comparison of MSE, MAE, and RMSE Across Different Prediction Models

H. Ammour al., PSDRE journal, Vol. 02 n. 01 (2025); ISSN 2992-040X, www.psdrej.net

Errors / Model	MSE	MAE	RMSE
random forest model	5080991.93	1216.6564	2254.1055
(xtrême Gradient Boosting	11160929.14	1641.1312	3340.7976
Decision Tree Model	6056277.98	1304.1840	2460.9506

The error metrics indicate that the Random Forest model performs the best among the three models for predicting solar energy production in our case study. In our comparison:

- The Random Forest model stands out with superior prediction performance. It has the lowest MSE at 5080,991.93, indicating that its predictions are, on average, closest to the actual values compared to the other models. Additionally, with an MAE of 1216.66, this model shows the smallest average absolute deviations from the actual values. The RMSE of 2254.11 further confirms its better performance with smaller prediction deviations.
- In comparison, the XGBoost (Extreme Gradient Boosting) model shows a higher MSE of 11160929.14, an MAE of 1641.13, and an RMSE of 3340.80, reflecting significantly lower performance compared to the Random Forest model.
- The Decision Tree model displays intermediate errors, with an MSE of 6056277.98, an MAE of 1304.18, and an RMSE of 2460.95, placing it between Random Forest and XGBoost in terms of prediction accuracy.

In conclusion, the Random Forest model provides the most accurate and reliable predictions, outperforming both the XGBoost and Decision Tree models across all performance indicators.

6. Conclusion

Effective real-time energy management is a crucial pillar in the transition to a more sustainable and resilient energy system. In this context, accurate forecasting of renewable energy production, particularly solar energy, as well as energy consumption, is of paramount importance. These forecasts enable the optimization of available resources, ensure the balance between supply and demand, and guarantee the stability of electrical grids.

One of the main characteristics of solar energy is its inherent variability, influenced by factors such as weather conditions, seasonality, and geographic location. To maximize its integration into the electrical grid and ensure its efficient management, it is essential to have reliable and accurate forecasting models. Similarly, anticipating energy consumption is crucial to avoid demand peaks and imbalances in electricity distribution.

In this regard, machine learning techniques offer powerful tools for energy time series modeling and forecasting. Among these techniques, Random Forest models, Decision Tree models, and XGBoost models stand out for their ability to handle complex datasets, identify nonlinear relationships, and provide accurate forecasts. These models have been widely used in the field of energy forecasting, demonstrating their effectiveness and robustness in various operational contexts.

References

1. Hatziargyriou, N. (2014). *Microgrids: Architectures and Control*. Wiley-IEEE Press.

2. Kouro, S., Buticchi, G., Palone, F., & Muñoz, J. (2020). Machine learning for energy management in microgrids: A review. *IEEE Transactions on Smart Grid*, *11*(2), 1341-1352.
3. Laraki, M.-H. (2019). Mémoire présenté à l'école de technologie supérieure comme exigence partielle à l'obtention de la maîtrise avec mémoire en génie, énergies renouvelables et efficacité énergétique. Montréal, Canada: École de technologie supérieure, Université du Québec.
4. Pamulapati, T., Cavus, M., Odigwe, I., Allahham, A., Walker, S., & Giaouris, D. (2023). A review of microgrid energy management strategies from the energy trilemma perspective. Energies, 16(2), 289.
5. Zahraoui, Y., Alhamrouni, I., Mekhilef, S., Basir Khan, M. R., Seyedmahmoudian, M., Stojcevski, A., & Horan, B. (2021). Energy management system in microgrids: A comprehensive review. Sustainability, 13(19), 10492.
6. Olama, M. M., Kuruganti, T. E., Nutaro, J. J., & Djouadi, S. M. (2017). Energy demand and management strategies for future smart grid. IEEE Power & Energy Society General Meeting.
7. Babu, A. N., Sudhakar, K., & Samykano, M. (2019). Optimal energy management strategy for solar PV and battery storage system. Sustainable Cities and Society, 44, 711-718.
8. Morvaj, B., Lugaric, L., & Krajcar, S. (2011). Demonstrating smart grid technology and management of energy demand through microgrid concept. Proceedings of the 2011 International Conference on IEEE Smart Grid Communications, 211-216.
9. Zhang, J., Wang, J., Wang, X., & Zhao, P. (2018). Prediction of Solar Power Generation and Energy Consumption Based on Machine Learning Algorithms. Sustainability, 10(8), 2678.
10. Gupta, H., Saini, L. M., & Srivastava, L. (2016). "Optimal Placement of Renewable Energy Resources in Microgrid Using Evolutionary Algorithms." IEEE Transactions on Smart Grid, 7(2), 1213-1221
11. Fathy, A., & Maleki, A. (2018). "A Hybrid Approach Based on Machine Learning for Forecasting Solar Energy Generation." Renewable Energy, 126, 299-311
12. Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. Stata Journal, 20(1), 3-29.
13. Georganos, S., Abdi, A. M., Tenenbaum, D. E., & Kalogirou, S. (2019). Machine learning in geography: A review focused on Random Forest. Applied Geography, 104, 102-112.
14. Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, *20*(1), 3-29.
15. Loh, W. Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 14-23.
16. Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 21(3), 660-674.
17. Kelley, K., & Raoul, G. (2020, 8 juin). Machine Learning : les 9 types d'algorithmes les plus pertinents en entreprise
18. Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies, 58, 308-324.
19. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 785–794.

20. Bouthevillain, K., & Mathis, A. (1995). Prévisions : mesures, erreurs et principaux résultats. Économie et Statistique, n°285-286, 89-100.
21. Thibout, C. (2019). La compétition mondiale de l'intelligence artificielle. Pouvoirs, n° 3, 131-142.
22. Russell, S., & Norvig, P. (2010). Intelligence artificielle: Fondements et applications - Pearson Education France.