Comparative Analysis of Machine Learning Models in Predictive Manners

Aymeric LEGROS[1] and Naveen JOHNSON VALLAVANATT[1]

[1] Data Science, EURECOM, France

Abstract

This project compares the performance of LASSO and k-Nearest Neighbors (kNN) regression models for predicting daily DC power generation from a solar plant, using weather conditions and lagged features. The study reveals that LASSO regression, particularly with hyperparameter tuning (LassoCV), generally outperforms kNN. However, both models struggle to achieve high accuracy, highlighting the need for improved feature engineering and a larger, more diverse dataset. The findings underscore the importance of careful model selection, hyperparameter optimization, and data quality in building accurate solar power generation forecasting models.

Index Terms: Solar power prediction, kNN regression, LASSO regression, Time series analysis, Feature selection, Model evaluation, Root Mean Square Error (RMSE), Renewable energy forecasting, Day-ahead forecasting.

1 Introduction

This report explores the application of LASSO regression, a linear regression technique with regularization, and kNN regression, a non-parametric method, to forecast daily DC power output from a solar plant. The study utilizes a dataset containing historical solar power generation data and corresponding weather information, including temperature and irradiation. By leveraging these features and incorporating lagged variables, the models aim to capture temporal dependencies and patterns in the data, leading to improved prediction accuracy.

The project evaluates the performance of both univariate and multivariate LASSO regression models, as well as kNN models with varying numbers of neighbors. The univariate models rely solely on historical power data, while the multivariate models incorporate additional weather features. The impact of using hourly data versus daily averages on prediction accuracy is also investigated.

This report is structured as follows: Section 2 provides a background and literature review on solar power prediction and discusses various approaches explored in previous studies, including statistical methods, machine learning techniques, and deep learning models. Works employing LASSO and kNN for one-day-ahead predictions are also explored. Section 3 details the methodology employed, including data preprocessing, feature engineering, model development, and evaluation metrics. Section 4 presents the results and discussion, comparing the performance of different models and analyzing feature importance and suggestions for future work. Finally, Section 5 concludes the report with a summary of findings and challenges.

2 Background and Related Work

Accurate prediction of solar power generation is essential for:

- **Efficient Grid Management:** Balancing energy supply and demand.
- **Renewable Energy Integration:** Maximizing the use of solar power.
- **Economic Benefits:** Optimizing energy trading and reducing reliance on fossil fuels.

Various approaches, including statistical methods and machine learning techniques, have been explored for solar power prediction.

Zafarani et al. [1] delved into the significance of weather data in solar power prediction by analyzing its impact on photovoltaic power generation forecasting accuracy. They identified key weather parameters influencing solar power output, emphasizing the importance of incorporating weather information for accurate predictions. The study employed various machine learning models, including support vector machines and artificial neural networks, to predict solar power generation and assessed the relative importance of different weather features.

Alanazi et al. introduced a day-ahead solar forecasting model leveraging multi-level solar measurements, utilizing a nonlinear autoregressive with exogenous input (NARX) model. By incorporating solar measurements from diverse locations, including customer, feeder, and substation levels, their model aimed to improve the accuracy of solar photovoltaic (PV) generation forecasts. The study compared the performance of the proposed model with two-level and single-level studies, demonstrating the advantages of incorporating multi-level measurements for enhanced forecasting accuracy.

Perera et al. [3] focused on day-ahead regional solar power forecasting by proposing two deep-learning-based methods that effectively leverage aggregated and individual power generation time series with weather data. They introduced two hierarchical temporal convolutional neural network (HTCNN) architectures and two strategies to adapt HTCNNs for regional solar power forecasting. Their work involved evaluating the proposed approaches using a large dataset collected over a year from 101 locations across Western Australia to provide a day-ahead forecast at an hourly time resolution. The results demonstrated the effectiveness of HTCNNs in reducing regional forecast errors and requiring fewer individually trained networks compared to alternative methods.

Luo et al. [4] tackled the limitations of offline learning in deep learning models by presenting an Adaptive LSTM (AD-LSTM) framework for day-ahead photovoltaic power generation forecasting. The AD-LSTM model dynamically learns from new data while preserving knowledge from historical data, making it adaptable to changes in the PV system and resilient to concept drift. The study evaluated the AD-LSTM model using multiple datasets from PV systems and demonstrated its superior forecasting accuracy compared to offline LSTM and other traditional machine learning and statistical models.

Dao et al. [5] investigated the application of ensemble methods for enhancing short-term and medium-term solar and photovoltaic power prediction. Their research encompassed a hierarchical structure for solar radiation prediction, utilizing machine learning techniques for data clustering before applying ensemble methods. They also explored different time series models and the integration of weather forecast services to im-

prove the prediction accuracy. The implementation of these ensemble methods on a low-cost Raspberry Pi platform demonstrated the feasibility of their approach in real-world scenarios.

Tang et al. [6] proposed a Least Absolute Shrinkage and Selection Operator (LASSO)-based forecasting model for solar power generation based on historical weather data. Their approach involved developing an algorithm that maximizes Kendall's tau coefficient to estimate prediction model coefficients, leveraging LASSO's variable selection capability to balance prediction accuracy and model complexity. The study evaluated the LASSO-based scheme with real-world datasets and found it to outperform existing methods, demonstrating its effectiveness in solar power generation forecasting.

Several other studies have explored day-ahead PV power forecasting using various methodologies. Gigoni et al. [7] conducted an extensive comparison of simple and sophisticated forecasting methodologies across 32 photovoltaic plants to evaluate the impact of weather conditions and forecasts on prediction accuracy. Conte et al. [7] focused on day-ahead and intra-day planning of integrated battery energy storage systems (BESS) and photovoltaic systems for frequency regulation, taking into account uncertainties in photovoltaic generation and frequency dynamics. Jiang et al. [9] proposed a day-ahead PV power forecasting method based on multiple seasonal-trend decomposition using LOESS (MSTL) and temporal fusion transformers (TFT), achieving improved accuracy compared to existing methods on a desert knowledge Australia solar centre dataset.

On an other field than power prediction, Shengtao Gao [10] published a study on trend-based stock price prediction method that employs the K-nearest neighbors (kNN) algorithm for trend forecasting. Experiments were conducted using a historical stock price dataset, and the prediction performance was evaluated. Evidences suggests that, in relation to accuracy in stock price prediction, the trend-based kNN algorithm exhibits superior performance over conventional machine learning approaches. In addition, the impact of prediction time span on model performance was investigated. The findings suggest that the trend-based kNN algorithm exhibits clear advantages when dealing with predictions over larger time spans.

The present study contributes to this body of research by focusing on univariate and multivariate day-ahead photovoltaic DC power prediction using LASSO and k-nearest neighbors (kNN) algorithms. While these algorithms have been explored in various contexts, their application to DC power prediction and comparison in a univariate and multivariate setting with one-day lag is novel. This research aims to provide insights into the comparative performance of LASSO and kNN for this specific prediction task, potentially offering valuable information for power system operators and renewable energy stakeholders.

3 Methodology

This section outlines the methodological framework adopted for this study on predicting solar power generation. The approach leverages machine learning techniques, specifically LASSO Regression with and without cross-validation, and k-Nearest Neighbors (kNN) with and without hyperparameter tuning, to model the relationship between weather conditions, lagged features, and DC power output.

3.1 Data

The project utilizes two datasets sourced from the "Solar Power Generation Data" repository available in Kaggle [11]:

1. **Generation Data:** This dataset encompasses records of DC power, AC power, daily yield, and total yield, all captured at 15-minute intervals across a span of 34 days.
2. **Weather Sensor Data:** This dataset provides readings for ambient temperature, module temperature, and irradiation, also recorded at 15-minute intervals over the same 34-day period.

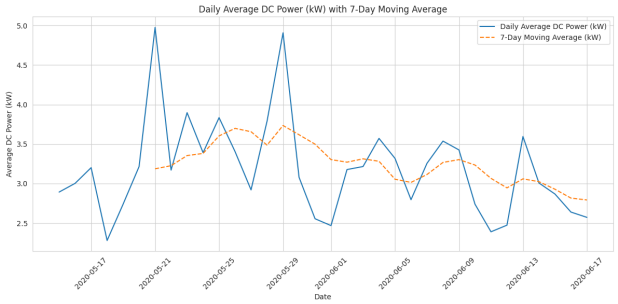

Figure 1. Average DC Power by Date and 7-Day Moving Average

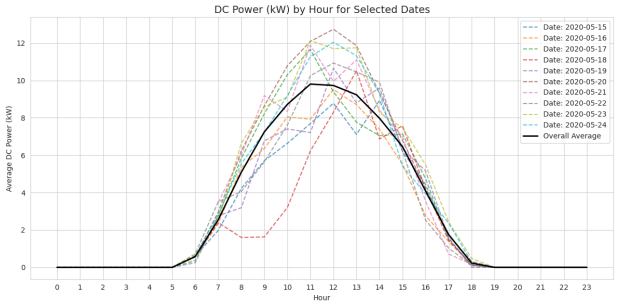

Figure 2. DC Power by Hour for a Selection of Dates and Overall Mean of the Entire Data

A brief exploratory analysis of the daily and hourly DC power generation patterns, visualized in Figure 1 and Figure 2, provide insights into the typical power output behavior and the influence of daily and hourly cycles.

3.2 Data Preprocessing

Prior to model training and analysis, the raw datasets underwent a series of preprocessing steps:

1. **Data Loading and Resampling:** Both datasets were loaded into the Python environment, with the date-time columns being converted into a consistent and analysis-friendly format. The initial 15-minute frequency of data collection was resampled into two different datasets with both daily and hourly aggregations. This provided flexibility in exploring models sensitive to different temporal resolutions.
2. **Merging Datasets:** The generation and weather data were merged into a unified dataset based on their corresponding date-time stamps. This ensured that weather features and power generation values were properly aligned for each observation.

3. **Feature Engineering:** Lagged features, representing past values of DC power and weather variables, were generated. These lagged features aimed to capture potential time-dependent patterns and enhance the predictive power of the models.
4. **Imputation:** The KNN Imputer was employed to address any remaining missing values in the dataset. This method leverages the values of neighboring data points to estimate and fill in missing entries, preserving data integrity for model training.

It is important to note that data normalization was skipped, despite it usually being a crucial step in machine learning pipelines, to preserve the original scale of the data and provide a clearer sense of the models' performance metrics.

3.3 Model

This study explored both linear and non-linear modeling approaches, with variations in hyperparameter settings, to comprehensively evaluate their effectiveness in predicting DC power generation:

1. **LASSO Regression:** A linear regression technique that incorporates L1 regularization, adding a penalty to model complexity and shrinking coefficients of less important features. This model was implemented with a default regularization parameter.
2. **LassoCV (LASSO Regression with Cross-Validation):** Extends LASSO regression by using cross-validation to find the optimal regularization parameter (λ). LassoCV tested a range of λ values from 0.001 to 1 to prevent overfitting and select relevant features. Time series cross-validation (TimeSeriesSplit), implementation as seen in Figure 3 was specifically employed to respect the temporal order of the data during model evaluation.
3. **kNN (k-Nearest Neighbors):** A non-parametric method that predicts an output based on the values of its k nearest neighbors in the feature space. This model was implemented using hyperparameter tuning for an automated selection of the best k neighbors. The model explored k values from 3 to 6 to minimize prediction error.
4. **kNN with Cross-Validation:** Improves upon the basic kNN by using cross-validation on top of hyperparameter tuning to determine the optimal number of neighbors (k). Similar to LassoCV, time series cross-validation was used to maintain temporal consistency during model assessment.

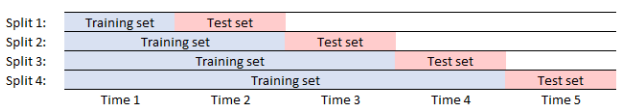

Figure 3. Expanding Window in scikit-learn's TimeSeriesSplit

All models were trained using an 80% split of the data and tested on the remaining 20%, with a temporally consistent division without shuffling to preserve the continuity of the time series and prevent data leakage.

3.4 Experiments

Here, and in the proceeding sections we abbreviate several terms: D for DC power, A for ambient temperature, M for module temperature, and I for irradiation. To evaluate the performance of LASSO and kNN under various scenarios, the following experiments were designed:

1. **Univariate One-day-ahead Prediction with 1 Lag:** This experiment focused on predicting the current day's average DC power solely based on the previous day's average DC power.

$$\bar{D_0} \xrightarrow{\text{Day 1}} \bar{D_1}, \quad \bar{D_1} \xrightarrow{\text{Day 2}} \bar{D_2}$$

$$\bar{D_{25}} \xrightarrow{\text{Day 26}} \bar{D_{26}}, \quad \bar{D_{26}} \xrightarrow{\text{Day 27}} \bar{D_{27}}, \quad \bar{D_{27}} \xrightarrow{\text{Day 28}} \bar{D_{28}}$$

$$\bar{D_{32}} \xrightarrow{\text{Day 33}} \bar{D_{33}}$$

2. **Multivariate (Multiple Features) One-day-ahead Prediction with 1 Lag:** This experiment expanded upon the previous one by incorporating additional lagged features, including daily average ambient temperature, module temperature, and irradiation. The goal was to assess whether the inclusion of these extra features improved predictive accuracy.

$$\bar{D_0}\ \bar{I_0}\ \bar{A_0}\ \bar{M_0} \xrightarrow{\text{Day 1}} \bar{D_1}, \quad \bar{D_1}\ \bar{I_1}\ \bar{A_1}\ \bar{M_1} \xrightarrow{\text{Day 2}} \bar{D_2}$$

$$\bar{D_{25}}\ \bar{I_{25}}\ \bar{A_{25}}\ \bar{M_{25}} \xrightarrow{\text{Day 26}} \bar{D_{26}}, \quad \bar{D_{26}}\ \bar{I_{26}}\ \bar{A_{26}}\ \bar{M_{26}} \xrightarrow{\text{Day 27}} \bar{D_{27}}, \quad \bar{D_{27}}\ \bar{I_{27}}\ \bar{A_{27}}\ \bar{M_{27}} \xrightarrow{\text{Day 28}} \bar{D_{28}}$$

$$\bar{D_{32}}\ \bar{I_{32}}\ \bar{A_{32}}\ \bar{M_{32}} \xrightarrow{\text{Day 33}} \bar{D_{33}}$$

3. **Multivariate (Hourly Train Data as Features) One-day-ahead Prediction with 1 Lag:** Recognizing the potential value of higher temporal granularity, this experiment used the previous day hourly weather data and lagged features to predict current day average DC power. Various combinations of features were explored through feature selection.

$$D_0^1 \ldots D_0^{24} \ldots I_0^1 \ldots I_0^{24} \xrightarrow{\text{Day 1}} \bar{D_1}, \quad D_1^1 \ldots D_1^{24} \ldots I_1^1 \ldots I_1^{24} \xrightarrow{\text{Day 2}} \bar{D_2}$$

$$D_{25}^1 \ldots D_{25}^{24} \ldots I_{25}^1 \ldots I_{25}^{24} \xrightarrow{\text{Day 26}} \bar{D_{26}}, \quad D_{26}^1 \ldots D_{26}^{24} \ldots I_{26}^1 \ldots I_{26}^{24} \xrightarrow{\text{Day 27}} \bar{D_{27}}, \quad D_{27}^1 \ldots D_{27}^{24} \ldots I_{27}^1 \ldots I_{27}^{24} \xrightarrow{\text{Day 28}} \bar{D_{28}}$$

$$D_{32}^1 \ldots D_{32}^{24} \ldots I_{32}^1 \ldots I_{32}^{24} \xrightarrow{\text{Day 33}} \bar{D_{33}}$$

3.5 Evaluation

Model performance was rigorously evaluated using the Root Mean Squared Error (RMSE) metric. RMSE quantifies the aver-

age magnitude of prediction errors, providing a measure of how well the model's predictions align with the actual DC power values. Lower RMSE values indicate better predictive accuracy.

$$RMSE : \sqrt{\frac{1}{6}\sum_{i=1}^{6}\left[(y_{test})_i - Model.Predict(x_{test})_i\right]^2}$$

A comparative analysis of RMSE scores across all experiments helped determine the most effective modeling approach for this specific problem. The analysis considered factors such as feature set selection, temporal resolution of the data, and the influence of hyperparameter tuning on each model's predictive accuracy.

4 Results and Discussion

This section discusses the findings of the experiments, analyzing the strengths and weaknesses of the proposed models. The analysis focuses on the following aspects:

- **Effectiveness and comparison of the models:** Each model's performance at different frequencies and the reasons for their effectiveness.
- **Limitations of the approach:** Potential limitations of each model.
- **Future work:** Future directions that could be explored to improve the performance of the models.

4.1 Effectiveness and Comparison of the Models

Initially, we focused on predicting daily average DC power using both univariate (only lagged DC power) and multivariate (including lagged ambient temperature, module temperature, and irradiation) setups.

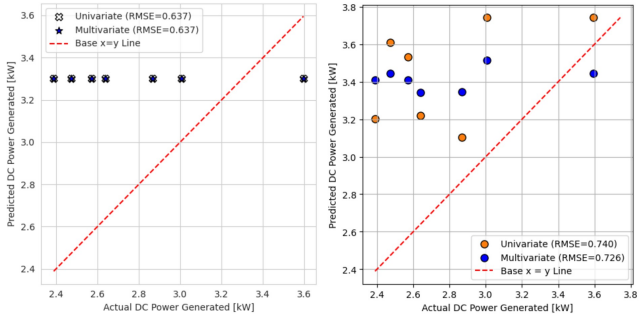

Figure 4. Univariate with only DC Power and Multivariate: Actual vs Predicted for LASSO (left) and kNN (right)

As shown in Figure 4, for the daily prediction frequency, the LASSO model with feature sets (D) and (D, A, I, M) both achieved an RMSE of 0.637. The kNN model performed slightly worse, with RMSE values of 0.740 for feature set (D) and 0.726 for feature set (D, A, I, M). This indicates that for daily predictions, the LASSO model is more effective. It's important to note that in both univariate and multivariate settings, the LASSO model, which utilized the default alpha (λ) value of 1.0, exhibited a particular behavior. When alpha is set to 1.0, the regularization penalty is high, leading to the coefficient for DC power being shrunk to zero for the univariate model and all the other features including DC power in the multivariate model. This effectively nullifies the influence of these features on the prediction. In such a scenario, the LASSO model relies solely on the intercept term to make predictions. Consequently, regardless of the input, the model will always predict the average DC power observed during training. This behavior explains the consistent RMSE value of 0.637 for LASSO in both the univariate and multivariate daily predictions. While LASSO technically outperforms kNN in these cases, its predictive power is limited by the high regularization strength, rendering it equivalent to a simple average predictor. However, the high RMSE values for both models highlight the lack of precision we encountered. Therefore, we transitioned to hourly training data to aim for an RMSE of 0.5 or lower, and perform hyperparamter tuning to achieve better regularization.

On an hourly basis, we examined which features were most relevant and how well each algorithm could capture the trend. It was found that Irradiation and DC power were the most informative features, likely due to their high correlation, as depicted in Figure 5.

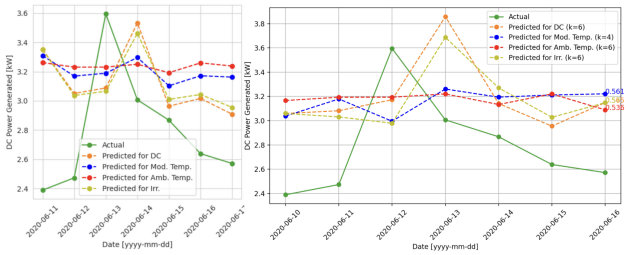

Figure 5. Trends for univariate prediction with time series cross-validation on an hourly training basis: LASSO (left) and kNN (right)

However, even with time series cross-validation and hyperparameter tuning in multivariate settings with DC power as a reference feature, we were unable to improve the models' ability to capture the intra-day variations in DC power. This is illustrated in Figure 6.

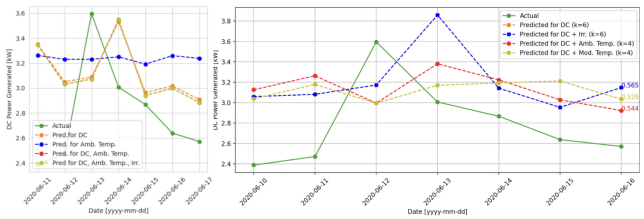

Figure 6. Trends for multivariate with feature set combinations: LASSO (left) and kNN (right)

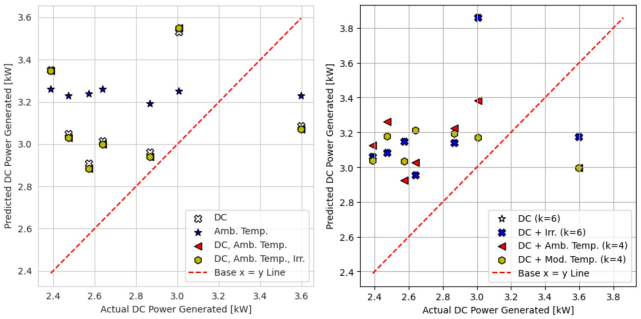

Figure 7. Multivariate with feature set combinations: Actual vs Predicted for LASSO (left) and kNN (right)

As shown in Figure 7, both models tend to overpredict on average, which suggests a potential area for future improvement.

A detailed examination of the LASSO coefficient table (Table 1) indicates that most coefficients, even during peak power generation hours (6 AM to 6 PM), are zero. This sparsity highlights

Table 1
LASSO coefficient selection with feature combinations

β_h	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
D	0	0	0	0	0	0
A	0	0	0	0	0	0
D, A	0,0	0,0	0,0	0,0	0,0	0,0
D, A, I	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0
β_h	$h = 7$	$h = 8$	$h = 9$	$h = 10$	$h = 11$	$h = 12$
D	0	0	0	0	0	**0.04**
A	0	0	0	0	0	0
D, A	0,0	0,0	0,0	0,0	0,0	**0.04**,0
D, A, I	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	**0.04**,0,0
β_h	$h = 13$	$h = 14$	$h = 15$	$h = 16$	$h = 17$	$h = 18$
D	**0.08**	0	0	0	0	0
A	**0.01**	0	0	0	0	0
D, A	**0.08**,0	0,0	0,0	0,0	0,0	0,0
D, A, I	**0.08**,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0
β_h	$h = 19$	$h = 20$	$h = 21$	$h = 22$	$h = 23$	$h = 24$
D	0	0	0	0	0	0
A	**0.02**	0	0	0	0	0
D, A	0,0	0,0	0,0	0,0	0,0	0,0
D, A, I	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0

LASSO's tendency to select only the most critical features, effectively ignoring others. Notably, specific hours such as hour 12 and 13 are used by LASSO, reinforcing the model's focus on particular time slots for prediction. Furthermore, analysis of the alpha vs. RMSE graph (Figure 9) demonstrates that adding more than two feature combinations results in similar outcomes, effectively removing those additional features from being used in training. This behavior underscores the importance of feature selection in optimizing model performance without overfitting.

4.2 Limitations of the Approach

The primary limitation we encountered stemmed from the dataset itself. The high volatility inherent in daily power output, combined with the relatively small sample size (34 days), likely hindered the models' ability to discern clear patterns.

This issue also affected the cross-validation step, as the small number of points in the test set (only 7) restricted the applicability of kNN for a higher number of neighbors, as shown in Figure 8.

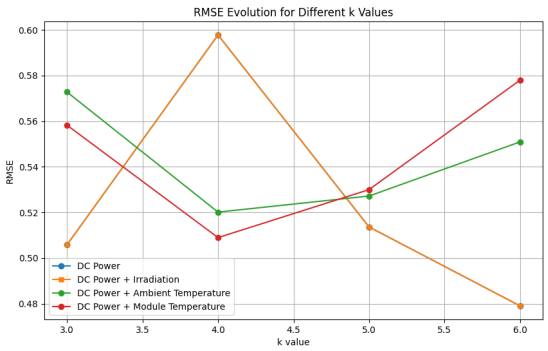

Figure 8. Evolution of RMSE as a function of the number of neighbors k in the kNN algorithm for multivariate with feature set combinations

Increasing the test set size to ten points confirmed this theory, allowing kNN to achieve an RMSE of 0.49 using DC power and module temperature as combined features.

It may be noticed that LASSO algorithm seemed to not be sensitive to this issue.

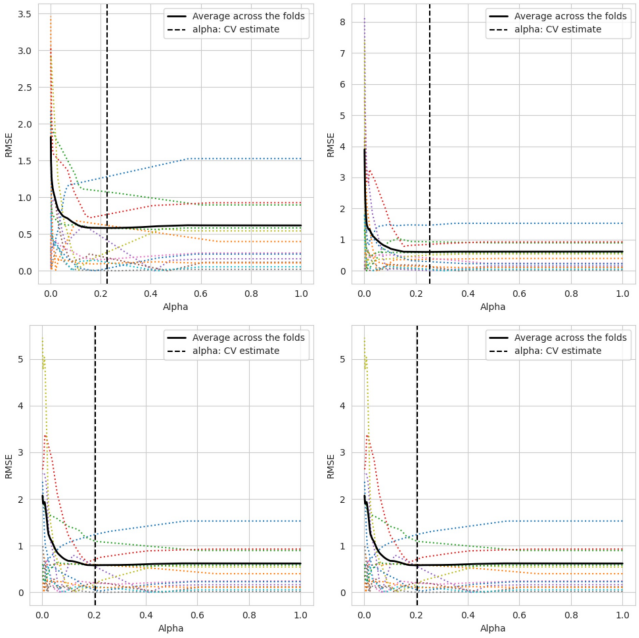

Figure 9. RMSE vs. regularization strength (λ) for LASSO models with varying feature sets. (Top left) DC power only, (top right) Ambient temperature only, (bottom left) DC power and ambient temperature, and (bottom right) DC power, ambient temperature, and irradiation

4.3 Future Work

Future research could explore the following directions:

- **Data processing:** Increase the dataset length or use a different dataset with more samples to improve the models' ability to capture trends and reduce the impact of volatility. A larger and more diverse dataset could enhance the robustness of the predictions.
- **Advanced Modeling Techniques:** Explore and implement advanced machine learning algorithms such as Random Forest, Gradient Boosting Machines, or Deep Learning models. These methods may provide better performance by capturing complex patterns in the data.
- **Feature Engineering:** Investigate additional features that could improve prediction accuracy, such as incorporating weather forecasts, historical power generation data, and real-time sensor data.
- **Optimization Techniques:** Further fine-tune hyperparameters using advanced optimization methods like Bayesian Optimization.

5 Conclusion

Based on the comparative analysis of machine learning models for predicting solar power generation from weather conditions and lagged features, several insights have emerged:

The study evaluated LASSO regression and kNN regression across different scenarios, including univariate and multivariate setups with daily and hourly data. Results indicated that LASSO regression generally outperformed kNN in both daily predictions and hourly predictions, even if kNN hit the lowest RMSE while sacrificing the trend catch. However, both models struggled to achieve desired accuracy levels, particularly in hourly

Table 2
Model Performance for Predicting Current Day Average DC Power

Model	Frequency	Features	k	λ	RMSE
Lasso	Day	D	-	1.0	0.637
Lasso	Day	D, A, I, M	-	1.0	0.637
LassoCV	Hour	D	-	0.225	0.542
LassoCV	Hour	A	-	0.253	0.593
LassoCV	**Hour**	D, A	-	**0.206**	**0.538**
LassoCV	Hour	D, A, I	-	0.206	0.538
kNN	Day	D	7	-	0.740
kNN	Day	D, A, I, M	7	-	0.726
kNN CV	Hour	D	6	-	0.565
kNN CV	Hour	A	6	-	0.536
kNN CV	Hour	I	4	-	0.570
kNN CV	Hour	M	6	-	0.561
kNN CV	Hour	D	6	-	0.565
kNN CV	Hour	D, A	4	-	0.544
kNN CV	**Hour**	D, M	**4**	-	**0.529**

predictions where the RMSE hovered around 0.5 or higher. The best performing models (Table 2) were LassoCV (RMSE of 0.538 using lagged DC power and ambient temperature) and kNN with cross-validation (RMSE of 0.529 using lagged DC power and module temperature).

Challenges such as dataset volatility and limited sample size impacted model performance, particularly in capturing finer temporal trends. Despite efforts in feature engineering and model tuning, including cross-validation and hyperparameter optimization, achieving precise predictions remained elusive.

In conclusion, while LASSO regression showed promise in daily predictions, enhancing data quality and model sophistication is necessary to improve forecasting accuracy for solar power generation. These improvements would also benefit kNN, as highlighted earlier. Future research should focus on acquiring a larger dataset, engineering domain-specific features, and exploring advanced time series models to address these challenges and further refine predictive capabilities in renewable energy forecasting.

Acknowledgements

This research received support during the ResProj course, from Professor Motonobu KANAGAWA and Ms. Parastoo PASHMCHI.

References

[1] Assessing the Utility of Weather Data for Photovoltaic Power Prediction (2018). https://arxiv.org/abs/1802.03913

[2] Day-Ahead Solar Forecasting Based on Multi-level Solar Measurements (2017). https://arxiv.org/abs/1710.03803

[3] Day-ahead regional solar power forecasting with hierarchical temporal convolutional neural networks using historical power generation and weather data (2024). https://arxiv.org/abs/2403.01653

[4] An Adaptive Deep Learning Framework for Day-ahead Forecasting of Photovoltaic Power Generation (2021). https://arxiv.org/abs/2109.13442

[5] Improving Solar and PV Power Prediction with Ensemble Methods (2020). https://arxiv.org/abs/2011.09950

[6] Solar Power Generation Forecasting With a LASSO-Based Approach (2018). https://ieeexplore.ieee.org/document/8306874

[7] Day-Ahead Hourly Forecasting of Power Generation From Photovoltaic Plants (2017). https://ieeexplore.ieee.org/document/8066360

[8] Day-Ahead and Intra-Day Planning of Integrated BESS-PV Systems providing Frequency Regulation (2021). https://arxiv.org/abs/2104.07352

[9] Day-Ahead PV Power Forecasting Based on MSTL-TFT (2023). https://arxiv.org/abs/2301.05911

[10] Trend-Based K-Nearest Neighbor Algorithm in Stock Price Prediction (2023). https://www.atlantis-press.com/proceedings/deca-23/125995001

[11] Ani Kannal. Solar Power Generation Data – Solar power generation and sensor data for two power plants. https://www.kaggle.com/datasets/anikannal/solar-power-generation-data. 2020.