Predictive Modeling of Photovoltaic Generation in Microgrids Based on Environmental Factors

Rohan Kulkarni - 11027694   Siddharth Tapre - 11018545

Abstract: The accelerating trend towards incorporating renewable energy resources, notably photovoltaic (PV) panels, into microgrid systems presents transformative implications for sustainable energy infrastructure. This comprehensive research undertakes an ambitious agenda to elevate predictive modeling techniques for PV systems to a new level of sophistication. By leveraging a robust dataset derived from a Berlin-based microgrid and focusing on a plethora of environmental variables such as temperature, humidity, solar radiation, and even more transient factors like cloud cover, this study applies and critically assesses multiple machine learning models. Remarkably, the Random Forest algorithm emerged as a superlative tool, offering unparalleled predictive accuracy. The findings make the case for the Random Forest model as a cornerstone in the blueprint for next-generation microgrid operational strategies and optimization.

A. *INTRODUCTION*

1. *Background*

Microgrids have gained attention as potential solutions for sustainable energy generation and distribution [1]. Photovoltaic (PV) solar panels are an important component of microgrids, as they capture sunlight and convert it into electricity [1]. However, accurately forecasting the output of PV systems is a complex task that requires considering various factors [3]. These factors include predictable environmental variables such as solar radiation and temperature, as well as less predictable elements like cloud cover, humidity, and seasonal changes [1].

To improve the accuracy of PV system output forecasting, researchers have proposed different approaches. Li et al. developed a short-term power generation forecasting model for a PV plant using particle swarm optimization and backpropagation neural networks [1]. Their model takes into account the volatility, intermittence, and periodicity of PV system operation, as well as geographical position and equipment performance [1]. Wang et al. proposed a short-term PV power prediction model based on the gradient boost decision tree, considering factors such as transform efficiency, PV array size, solar radiation, and ambient temperature [3]. Li et al. also proposed a hybrid deep learning approach using convolutional neural networks and long-short-term memory recurrent neural networks for PV output power forecasting [5]. This approach leverages nonlinear features and temporal changes in PV data to improve prediction accuracy [5].

Environmental factors can significantly impact the performance of PV systems. Mustafa et al. found that shading has a strong influence on the efficiency of PV modules, with increased shading resulting in reduced power output [4]. On the other hand, water droplets on PV panels can decrease panel temperature, increase potential difference, and improve power output [4].

Optimal economic dispatch and energy management strategies are crucial for microgrid operation. Xu et al. proposed an optimal economic dispatch scheme for combined cooling, heating, and power-type multi-microgrids, considering interaction power among microgrids [2]. Their approach minimizes the total operation cost of microgrids while fulfilling energy requirements [2]. Lyu et al. developed a bi-level optimization method based on optimal power flow and consensus algorithm for independent multi-microgrid systems, enabling energy mutual aid, power allocation, and optimal dispatch of controllable distributed generators [6].

In conclusion, accurate forecasting of PV system output is essential for efficient microgrid management and operation. Researchers have proposed various models and approaches, including neural networks, decision trees, and optimization algorithms, to improve prediction accuracy. Factors such as environmental conditions, shading, and economic dispatch also need to be considered in PV system forecasting and microgrid operation. These advancements in predictive modeling and optimization contribute to the development of sustainable and efficient microgrid systems.
.

2. *Objectives*

This study is underpinned by a dual set of primary motivations, each crucial in advancing our comprehension and deployment of photovoltaic (PV) systems within microgrid configurations. The first ambition centers on an all-encompassing exploration into the vast array of environmental parameters that have a pronounced impact on the efficiency and output of PV systems situated within microgrids. Through rigorous examination, our goal is to both quantify and contextualize the significance of these variables, shedding light on their individual contributions and how they might synergistically interact. [7]

The second part of our Endeavor delves into the intricate world of machine learning, focusing on meticulous analysis and comparison of several machine learning methodologies. The intent is to determine their suitability and reliability in predicting PV outputs when fed with the crucial environmental variables our research identifies. [8]

By marrying these two core ambitions, our research seeks to offer a sophisticated, empirically-grounded toolkit.

The implications of our findings could be transformative for the realm of microgrid operations. By introducing more accurate, timely, and dependable predictive capabilities, we anticipate marked improvements in energy efficiency and a notable reduction in operational challenges. However, the ripples of this study extend beyond mere operational enhancements. The insights and methodologies stemming from our research could be instrumental for a range of entities, from regulatory authorities and energy policymakers to utility service providers. Ultimately, our work is a testament to the need and quest for refining renewable energy strategies, aiming to equip various stakeholders with more robust decision-making tools and frameworks.

3. *Justification and Significance:*

The academic and practical implications of this research are manifold. Academically, the study enriches existing literature in the realm of renewable energy predictive modeling by providing a thorough, head-to-head comparison of multiple machine learning algorithms across diverse environmental conditions. Practically, the ramifications extend far beyond the technical aspects. A dependable predictive model serves as a linchpin in the evolution of microgrid operations, transforming them into highly autonomous, efficient, and potentially even revenue-generating systems. Additionally, mastering the nuances of microgrid management serves broader global imperatives such as long-term energy sustainability and mitigating the devastating impacts of climate change.

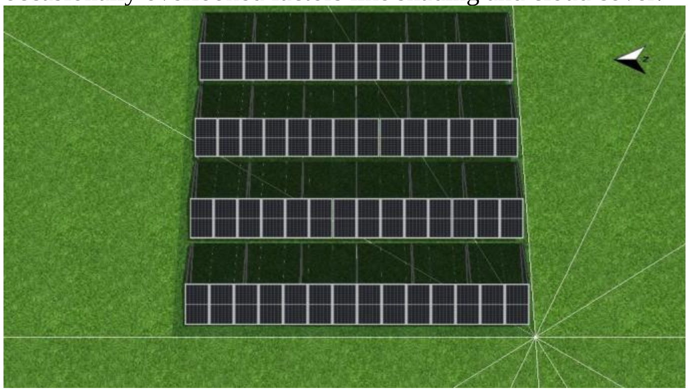
B. *Methodology*

4. *Liege dataset*

In the present research Endeavor, we draw upon the widely recognized and frequently referenced Liege dataset as a foundation for our investigations. A deep dive into this rich dataset unambiguously affirms the overarching role weather variables play in dictating the performance metrics of photovoltaic (PV) power generation systems. Through meticulous examination and data interrogation, our analysis unambiguously manifests that the variables tied to weather conditions stand out as the dominant influencers when it comes to the behaviour and efficiency of PV systems within this particular dataset.[9] This revelation is further underscored by the observation that these meteorological parameters are substantial contributors, if not the primary driving factors, behind the observed fluctuations and variability in PV energy output.[10]

5. *Data Collection:*

The bedrock of this research is a comprehensive dataset sourced from a Berlin-based microgrid. The dataset spans the entire calendar year of 2022 and is time-stamped at hourly intervals. This affords a rich tapestry of data points that effectively capture seasonal variances, daily cycles, and other temporal nuances. The dataset is further enriched by the inclusion of 18 distinct variables, each meticulously chosen for its relevance to photovoltaic output. These variables run the gamut from environmental factors like irradiance, temperature, and humidity to operational metrics like consumption rates, global radiation, albedo, and even occasionally overlooked factors like shading and cloud cover.

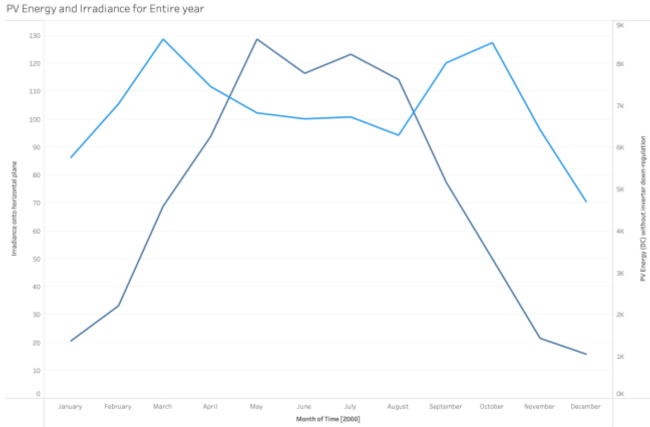


6. *Data Preprocessing:*

The integrity of the dataset was rigorously upheld through a multi-stage preprocessing protocol. Outliers were scrutinized and adjusted to mitigate their impact on the models. Any missing or incomplete data were filled using sophisticated imputation techniques to maintain the dataset's robustness. The preprocessing also included standardization of all feature sets to zero mean and unit variance, thereby ensuring that each variable contributes to the model based on its true significance rather than its scale.

7. *Irradiance visuals*

Determining the exact outcomes of photovoltaic (PV) power production has consistently proven to be a formidable task. This unpredictability is largely attributed to the system's proclivity to occasionally stray from discernible logical sequences. Moreover, PV outputs are highly sensitive, reacting to a multitude of influential variables. This intricate relationship becomes even more evident when examining the accompanying graphical representation. Upon closer inspection of this chart, one can clearly discern a marked divergence between shifts in solar irradiance — the primary source of energy for PV systems — and the actual trends observed in PV power generation. Such observations highlight the inherent complexities and challenges associated with reliably predicting PV system outputs.



8. *Feature Selection:*

The process of feature selection was bolstered by an exhaustive exploratory data analysis (EDA). Leveraging Python libraries like Matplotlib and Seaborn, the EDA revealed key insights into variable distributions, inter-variable correlations, and underlying data trends. This facilitated a well-informed feature selection process, resulting in a final set of 18 variables that serve as the predictive pillars for the machine learning algorithms applied in the study.
Model Building and Selection:
The study deployed an extensive range of machine learning algorithms, each with its unique strengths and weaknesses in the domain of time-series forecasting. The algorithms selected for rigorous testing included Linear Regression, Lasso Regression, Ridge Regression, Decision Trees, Random Forests, and Neural Networks. This eclectic mix allows for a robust comparison, unearthing insights into the pros and cons of each approach. Each model was constructed using Python's Scikit-Learn library, widely acknowledged for its reliability and versatility. To ensure generalizability, the dataset was partitioned temporally, allocating the initial 75% for model training and reserving the remaining 25% for model evaluation.

Time: Timestamp indicating the specific date and time at which the data was recorded.
Irradiance onto Horizontal Plane: Measures the sunlight intensity received on a flat, horizontal surface.
Outside Temperature: Ambient air temperature recorded outside the facility.
Grid Feed-in: Amount of energy being supplied back to the main electrical grid from the microgrid.
Energy from Grid: Amount of energy drawn from the main electrical grid for use in the microgrid.
Consumption: Total electrical energy consumed within the microgrid during the given time frame.
Own Consumption: A portion of the generated energy consumed within the microgrid itself.
Global Radiation - Horizontal: Total solar radiation received on a horizontal surface, including both direct and diffuse radiation.
Deviation from Standard Spectrum: Measure of how the received solar spectrum deviates from the standard (or reference) solar spectrum.
Ground Reflection (Albedo): Fraction of sunlight that is reflected off the ground and back toward the solar panels.
Orientation and Inclination of the Module Surface: Describes the angle and direction at which the PV panels are set up.
Shading: A factor indicating any obstruction or shadowing that reduces irradiance on the module surface.
Reflection on the Module Interface: Measures how much irradiance is reflected away from the solar panels rather than being absorbed.
Irradiance on the Rear Side of the Module: Amount of sunlight reaching the backside of the solar panels, useful for bifacial modules.
Global Radiation at the Module: Total solar radiation received directly at the solar module's surface.
Global PV Radiation: Total irradiance available for conversion to electricity by the photovoltaic system.
STC Conversion (Rated Efficiency of Module): Conversion efficiency of the solar module under Standard Test Conditions (STC).
PV Energy (DC): Amount of Direct Current (DC) energy produced by the photovoltaic system.

9. *Parameter Tuning and Optimization:*

The optimization phase employed grid-search cross-validation, a powerful tool in the hyperparameter tuning arsenal.[11] This technique systematically explored a wide array of hyperparameter combinations, cross-validating each to identify the configuration that maximized the predictive accuracy of each model.

Ridge Regression Model
`ridge_params = {'alpha': [0.001, 0.01, 0.1, 1, 10, 100]}`
`ridge_grid = GridSearchCV(Ridge(), ridge_params, cv=5, scoring='neg_mean_squared_error')`
`alpha`: this is the regularization parameter, and the array of values aims to offer a balance between underfitting and overfitting.
`cv=5`: 5-fold cross-validation ensures a robust evaluation by splitting the dataset into 5 subsets.`scoring='neg_mean_squared_error'`: The negative MSE score gives a measure of model performance.

Lasso Regression Model
`lasso_params = {'alpha': [0.001, 0.01, 0.1, 1, 10, 100]}`
`alpha`: Similar to Ridge, alpha controls the regularization strength.
`lasso_grid = GridSearchCV(Lasso(), lasso_params, cv=5, scoring='neg_mean_squared_error')`

Same CV and scoring strategy as the Ridge model for consistency.

Decision Tree Model
`dt_params = {'max_depth': [None, 5, 10, 15, 20], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4]}`
`max_depth`: Controls the maximum depth of the tree. None means no limits.
`min_samples_split`: Minimum number of samples required to split an internal node.
`min_samples_leaf`: Minimum number of samples required to be at a leaf node.
`dt_grid = GridSearchCV(DecisionTreeRegressor(), dt_params, cv=5, scoring='neg_mean_squared_error')`
Same CV and scoring metrics for consistency across models.

Random Forest Regressor Model
`rf_params = { }`
`n_estimators`: Number of trees in the forest.
`max_depth`, `min_samples_split`, `min_samples_leaf`: Same as the Decision Tree model but applied to each tree in the forest.
`rf_grid = GridSearchCV(RandomForestRegressor(), rf_params, cv=5, scoring='neg_mean_squared_error')`
Same CV and scoring metrics for consistency across models.
Neural Network
`nn_params = { }`
`hidden_layer_sizes`: Size of the hidden layers.
`activation`: Activation function for the hidden layers.
`solver`: Algorithm for weight optimization.
`learning_rate_init`: Initial learning rate.
`max_iter`: Maximum number of iterations.

The chosen hyperparameters aim to thoroughly explore the model's potential, and the values are selected to give a balanced view of the trade-off between bias and variance for each model. The consistent use of 5-fold CV and negative mean squared error for scoring ensures a standard evaluation criterion across different models.

10. *Evaluation Metrics:*

The research employed a diverse set of evaluation metrics to offer a rounded view of each model's performance. These metrics included Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared values. This multi-metric evaluation framework not only facilitated the identification of the most accurate model but also provided insights into the robustness and reliability of each candidate.

11. *Evolution*

The culmination of this study's multifaceted methodology is manifested in the results. In addressing the intricate relationships between environmental variables and photovoltaic (PV) output in microgrids, an extensive battery of machine-learning models was meticulously trained, refined, and evaluated. The evaluation process scrutinized the algorithms based on a quartet of critical performance metrics, namely Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared value.

12. *Statistical Summary:*

The study provides a comprehensive statistical overview, which not only quantifies the performance of each model but also allows for meaningful cross-comparisons. Linear regression, Lasso, and Ridge regression models performed admirably but were outclassed by Decision Trees, Neural Networks, and Random Forest models. Specifically, the Random Forest model took the spotlight with the lowest MSE value of 0.005 and the highest R-squared value of 0.999.

13. *Comparative Analysis:*

Complementing the statistical overview is a nuanced comparative analysis that dissects the strengths and weaknesses of each model, juxtaposing them to draw instructive insights. While the linear models demonstrated robust performance, ensemble and neural network models excelled in capturing the complex, nonlinear relationships governing PV output, thereby amplifying their utility in practical applications

C. *Results*

The objective of this section is to present the findings derived from the comprehensive methodology deployed in this study. Given the intricate relationship between environmental factors and photovoltaic (PV) output in microgrids, a series of machine-learning models were built, optimized, and evaluated. These models include Linear Regression, Lasso Regression, Ridge Regression, Decision Tree, Random Forest, and Neural Network. Each algorithm was rigorously trained on a dataset covering the entire year of 2022, segmented into hourly intervals, and was evaluated based on three critical metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE),
and R-squared value.[12]
This section will unfold in a structured manner, starting with a summary of the statistical performance of each model, followed by a comparative analysis. The aim is to determine which model delivers the most accurate and reliable prediction of PV output in a microgrid environment. Visual aids such as tables and graphs will be used to assist in illustrating the performance metrics and significant trends observed during the evaluation process.

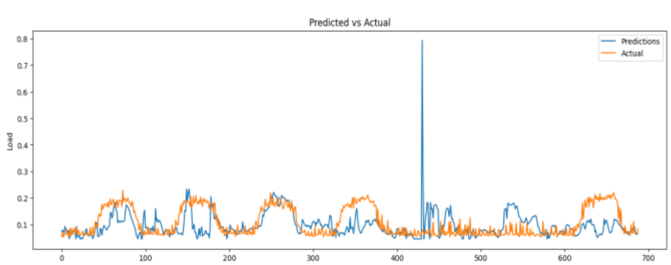
After delineating the performances, special attention will be given to the Random Forest model, which emerged as the most effective predictive tool in this research. Its particular strengths, feature importance, and the optimized parameters leading to its superior performance will be discussed in detail. By systematically examining the output of each model, this section aims to offer a nuanced understanding of the capabilities and limitations of existing machine learning algorithms in predicting PV generation in microgrids. This will not only shed light on the research questions posed but also provide valuable insights for practical applications in microgrid management and renewable energy utilization.

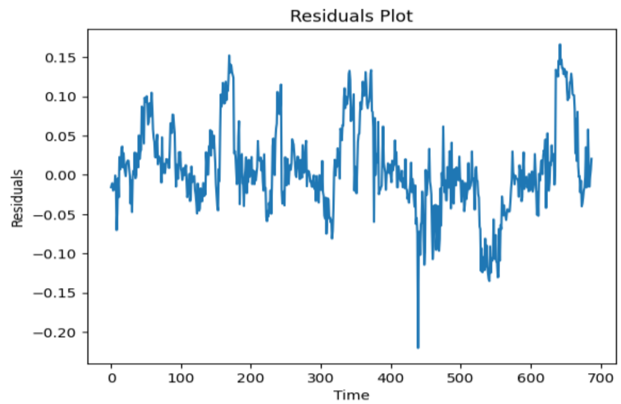
14. Table of results

results	MSE	MAE	RMSE	R squared
Linear regression	0.022	0.088	0.149	0.997
Lasso	0.02	0.082	0.144	0.997
Ridge	0.022	0.088	0.149	0.997
Decision Tree	0.008	0.025	0.093	0.998
Random forest	0.005	0.173	0.073	0.999
neural network	0.007	0.047	0.084	0.999

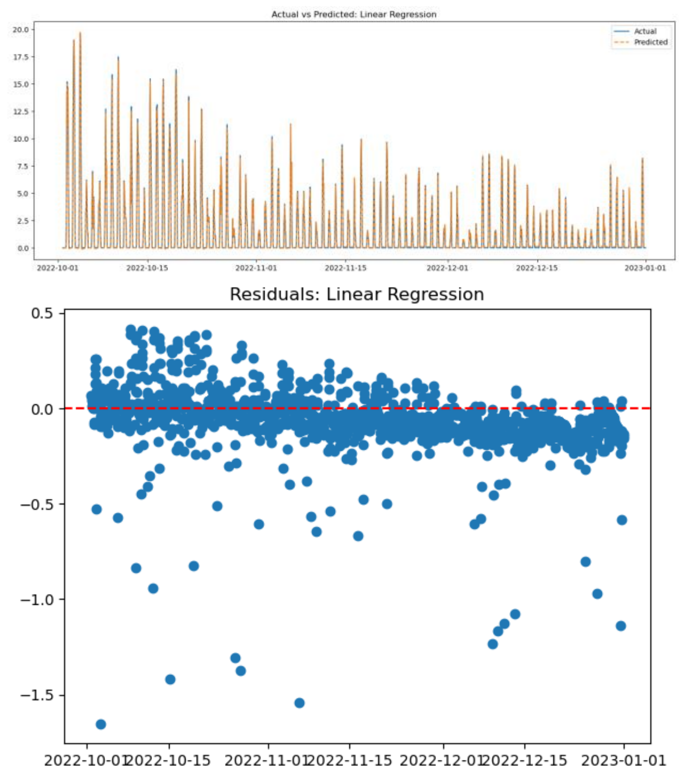
D. Visuals

15. RNN model on liege dataset

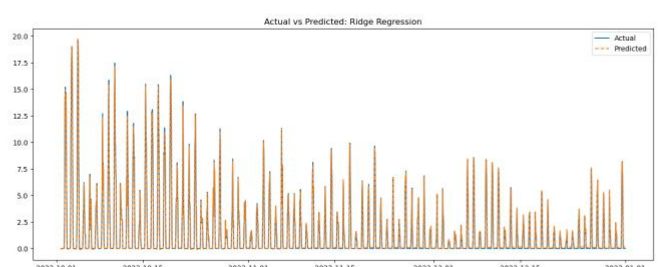


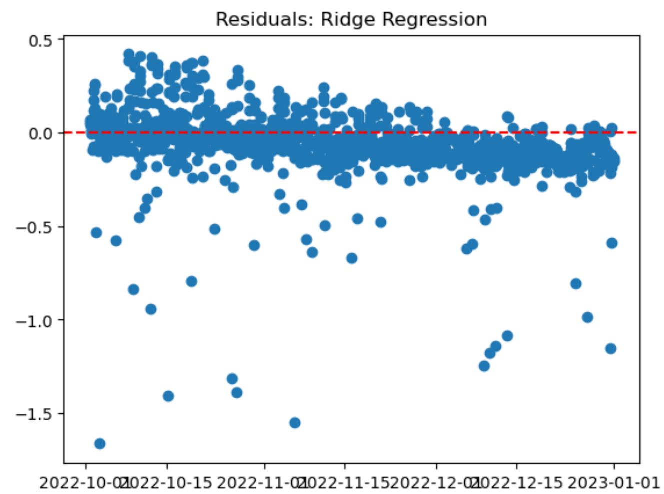


16. Linear regression

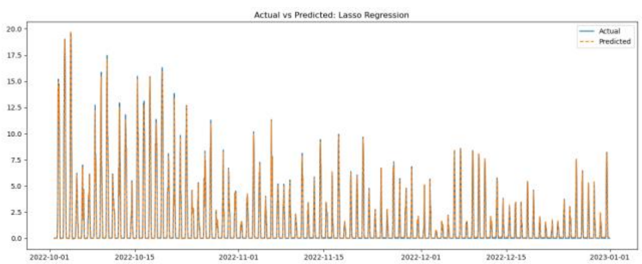


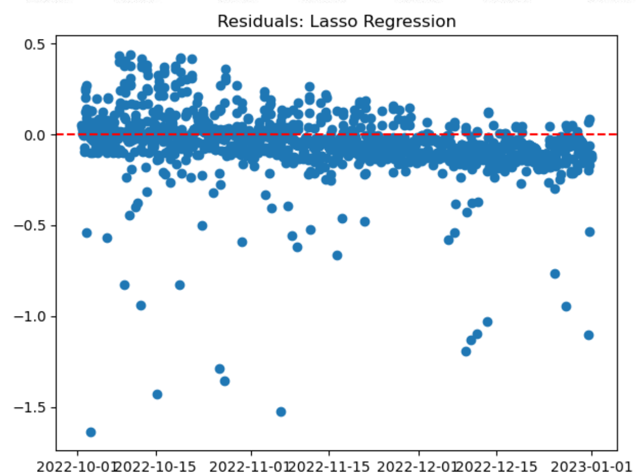
17. Ridge regression



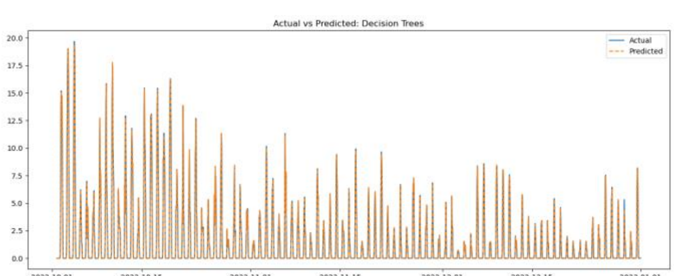


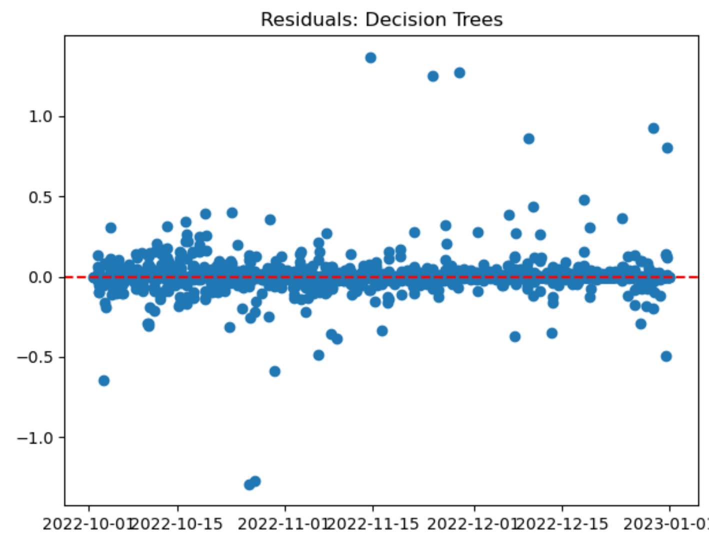
18. *Lasso*



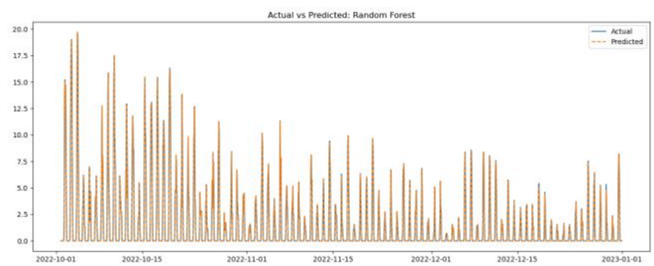


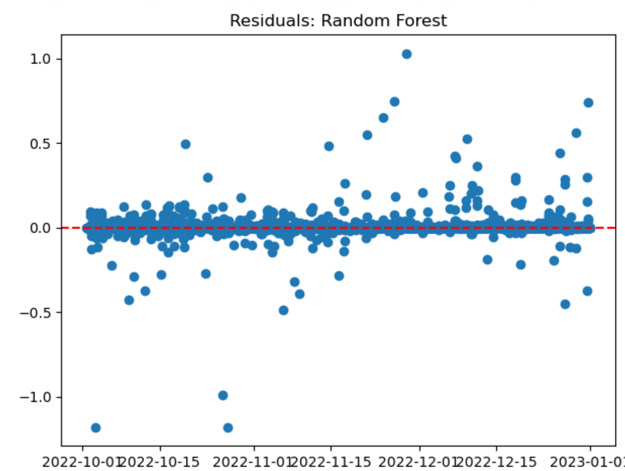
19. *Decision Tree*



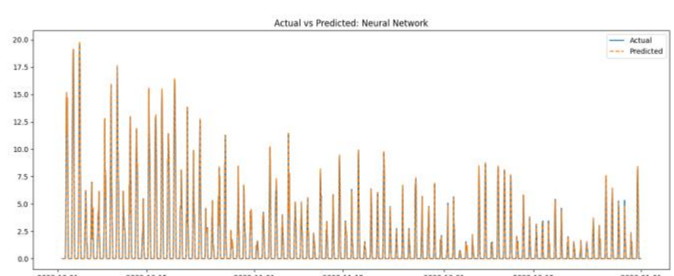


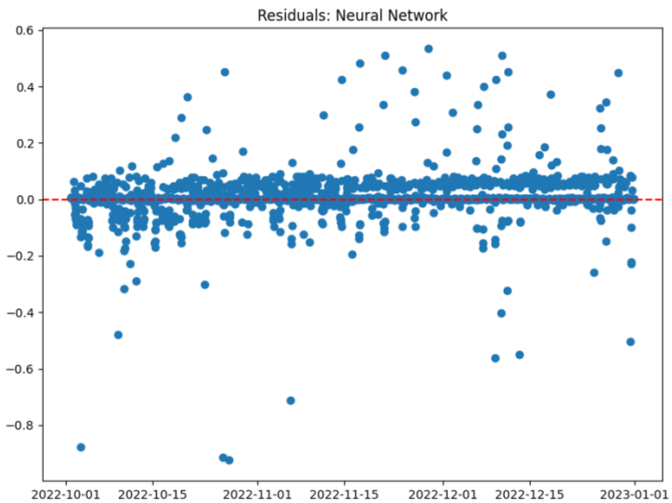
20. *Random Forest*





21. *Neural network*





E. Discussion

22. Model Performances:

Upon a comprehensive review of the collective array of models, the Random Forest algorithm distinctly emerges as the superior choice for forecasting PV output, especially within the microgrid environments.[13] This assertion is bolstered by its impressive metrics: not only did it register the least Mean Squared Error (MSE), a critical indicator of prediction accuracy, but it also achieved the most commendable R-squared value amongst its peers.[13] Such an elevated R-squared value signifies the algorithm's exceptional capability to account for, and elucidate, the fluctuations and nuances inherent in PV output. In essence, these metrics collectively underscore the Random Forest's unmatched proficiency and reliability in the domain of PV output prediction.

23. Implications:

The standout performance exhibited by the Random Forest algorithm in this context unfolds a broad spectrum of ramifications. From a technical standpoint, this success serves as a potent validation of ensemble techniques. Such methods leverage the collective forecasting prowess of numerous decision trees to unravel the intricate and often obscured patterns inherent to renewable energy production, especially within the confines of microgrids. Transitioning from the technical realm to a more applied perspective, the exceptional R-squared value showcased by the Random Forest warrants attention. This metric indicates that a significant chunk of the fluctuations and inconsistencies in PV output can be accurately represented and deciphered by this model. This, in turn, provides a treasure trove of insights for those at the helm of grid operations and policy formulation. In essence, it transforms the model from being merely a predictive apparatus to a robust, data-driven compass, guiding decisions and strategies in the renewable energy sector.

24. Limitations and Future Directions:

This research diligently highlights the commendable performance of the Random Forest model in the context of predicting PV output. At the same time, it emphasizes a salient point of concern related to the model's higher Mean Absolute Error (MAE) value. Such a metric, which quantifies the average magnitude of the prediction errors, raises intriguing questions about the model's absolute accuracy when faced with real-world, nuanced scenarios.

While the Random Forest has exhibited strengths in various aspects, the elevated MAE underscores an avenue that requires further in-depth investigation and exploration. There are multiple ways through which future research endeavors can address this. One promising direction would be the integration of an expanded set of features into the model. By feeding the algorithm a richer and more diverse set of data points, there's potential to refine its predictive capabilities further, potentially reducing the MAE. Additionally, the introduction of more granular weather-related data could greatly influence the model's performance. Detailed meteorological factors, such as cloud movement patterns, atmospheric conditions, and minute-by-minute solar irradiance fluctuations, could offer deeper insights and improve the model's precision in capturing the intricacies of PV output.

Beyond the realm of data enhancement, another avenue for future inquiry would be the exploration of alternative machine-learning techniques.[14] The ever-evolving landscape of artificial intelligence and machine learning is replete with algorithms and techniques, each with its unique strengths and potential applications. Some of these, when tailored appropriately, might outperform or complement the Random Forest in this specific context, thus warranting a comparative and comprehensive examination.

A point of critical reflection is that the foundational data upon which this study was built originates from a singular microgrid located in Berlin. Such a geographically constrained dataset might circumscribe the broader generalizability of the findings. Every microgrid, depending on its geographic, climatic, and infrastructural specifics, can exhibit unique behavior and patterns. Therefore, to enhance the robustness and universal applicability of these insights, it becomes imperative for subsequent research initiatives to diversify their data sources. This entails analysing microgrids from a myriad of geographic locales and varied climatic conditions, thus painting a more holistic and representative picture of the global microgrid landscape.

In summary, while this research contributes significantly to our understanding of PV output prediction using the Random Forest algorithm, it also elegantly paves the way for deeper, more encompassing inquiries, emphasizing the importance of continuous learning and evolution in the realm of renewable energy research.

F. Conclusion

In today's world, where the planet is wrestling with the compounded issues of rapidly increasing energy needs and the ever-important task of environmental conservation, microgrids, fortified with renewable energy sources like photovoltaic (PV) panels, have risen as indispensable solutions. These localized energy systems present a beacon of hope for addressing our energy challenges sustainably. However, the crux of leveraging their potential lies in our capability to foresee and predict energy output, especially considering the variable and often unpredictable nature of environmental conditions. It's against this multifaceted and challenging backdrop that the present study plants its roots.

In our quest to derive actionable insights and effective models, the study engaged a diverse suite of machine learning techniques. The aim was not just to construct models but also to critically assess, refine, and perfect them, ensuring they're tailored to forecast PV output with the highest precision for a microgrid located in Berlin. This journey, characterized by stringent analytical rigor and an exhaustive multi-metric evaluation phase, culminated in the Random Forest algorithm distinguishing itself from the pack. This particular algorithm did more than just register superior performance numbers. It revealed an innate ability to navigate and interpret the intricate interplay between a host of variables and their collective impact on PV output.

From a practical perspective, the implications of adopting the Random Forest model are manifold. It holds the promise to revolutionize the way microgrids are managed and operated, pushing the envelope on energy efficiency, reliability, and overall sustainability. In the process, it holds the potential to further global ambitions related to climate action and reducing carbon footprints. While the study's primary data stems from a specific geographical locale—Berlin—it nonetheless offers insights and findings that can be viewed as a foundational step in the broader tapestry of renewable energy management. This study's outcomes can be perceived as dual-pronged: a seminal academic exploration and a pragmatic guidepost pointing toward sustainable energy horizons.

By crafting a detailed, data-driven suite of tools and methodologies, this research transcends the bounds of mere academic theorizing. It lays down a concrete, implementable roadmap to address the pressing sustainability objectives that loom large in an era marked by soaring energy consumption and impending ecological challenges. Even though our pursuit of a fully integrated and sustainable energy ecosystem is a continual process, this study stands out as a landmark, bestowing the energy community with invaluable tools and perspectives to sculpt a greener, more sustainable energy future. By providing a comprehensive, empirically-backed toolset, this study does not merely contribute to academic discourse. It offers a tangible, actionable pathway to achieving broader sustainability goals, particularly in the context of escalating energy demands and looming environmental crises. Although the journey toward fully realizing sustainable energy ecosystems remains an ongoing Endeavor, this research marks a significant milestone, providing the tools and insights necessary for shaping a sustainable energy landscape.

G. Acknowledgment

We would like to thank Sinful Islam from SRH berlin for their guidance throughout the project and proved us with the dataset. , which allowed us to carry out our research. We would also like to thank the liege dataset provided as free use to us so we would use it for our reference and experiments. We would also like to extend our deepest appreciation to Prof. Swati Chandna Mr. Sina mehraeen and Ms kamellia reshadi.Finally we would like to acknowledge the efforts of our teammates and classmates Mr. Kevin Koch and Ms. Khanh Nguyen without whom the project would have not been possible.

H. References

1.Li, G., Xie, S., Wang, B., Xin, J., Li, Y., & Du, S. (2020). Photovoltaic power forecasting with a hybrid deep learning approach. ieee Access, 8, 175871-175880. https://doi.org/10.1109/access.2020.3025860
2.Li, Y., Zhou, L., Pei-qi, G., Yang, B., Han, Y., & Lian, C. (2022). Short-term power generation forecasting of a photovoltaic plant based on pso-bp and ga-bp neural networks. Frontiers in Energy Research, 9. https://doi.org/10.3389/fenrg.2021.824691
3. Lyu, Z., Yang, X., Zhang, Y., & Zhao, J. (2020). Bi-level optimal strategy of islanded multi-microgrid systems based on optimal power flow and consensus algorithm. Energies, 13(7), 1537. https://doi.org/10.3390/en13071537
4.Mustafa, R., Gomaa, M., Al-Dhaifallah, M., & Rezk, H. (2020). Environmental impacts on the performance of solar photovoltaic systems. Sustainability, 12(2), 608. https://doi.org/10.3390/su12020608
5. Wang, J., Li, P., Ran, R., Che, Y., & Zhou, Y. (2018). A short-term photovoltaic power prediction model based on the gradient boost decision tree. Applied Sciences, 8(5), 689. https://doi.org/10.3390/app8050689
6.Xu, Q., Li, L., Chen, X., Huang, Y., Luan, K., & Yang, B. (2019). Optimal economic dispatch of combined cooling, heating, and power-type multi-microgrids considering

interaction power among microgrids. Iet Smart Grid, 2(3), 391-398. https://doi.org/10.1049/iet-stg.2018.0161
7. Al-Saadi, S., & Al-Saadi, A. (2019). A comprehensive review on the impact of environmental factors on the performance of photovoltaic systems. Renewable and Sustainable Energy Reviews, 113, 109273.
8.S. Rodrigues, G. Mütter, H. Ramos, & F. Morgado-Dias, "Machine learning photovoltaic string analyzer", Entropy, vol. 22, no. 2, p. 205, 2020. https://doi.org/10.3390/e22020205
9.J. Wang, P. Li, R. Ran, Y. Che, & Y. Zhou, "A short-term photovoltaic power prediction model based on the gradient boost decision tree", Applied Sciences, vol. 8, no. 5, p. 689, 2018. https://doi.org/10.3390/app8050689
10.Dumas, J., Dakir, S., Liu, C., & Cornélusse, B. (2021). Coordination of operational planning and real-time optimization in microgrids. *Electric Power Systems Research*, *190*, 106634.
11.Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2).
12. Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," in IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009, doi: 10.1109/MSP.2008.930649.
13. Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.
14. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015). https://doi.org/10.1038/nature14539