A Comparative Analysis of Machine Learning Techniques for Solar Power Forecasting

Karan Sonawane

x22167005

Abstract

Solar power generation relies heavily on unpredictable environmental factors, making accurate forecasting essential for grid integration. This analysis evaluates machine learning techniques for forecasting photovoltaic (PV) power output using multivariate weather data from 12 northern hemisphere sites, without solar irradiance data that can have significant measurement errors. The methods examined include XGBoost Regression, Random Forest Regression, LightGBM, and an ensemble Stacking Regressor combining the models. The results demonstrate the Stacking Regressor achieves the best performance with an RMSE of 0.1370 and R-squared of 0.637 on the test set by leveraging the strengths of the component models. The data preprocessing, feature engineering, and ensemble architecture optimization were critical to maximizing accuracy. Overall, the research provides empirical evidence that advanced machine learning, specifically a tailored stacking ensemble approach, can effectively forecast solar PV power utilizing meteorological data as inputs without needing irradiation values that require additional equipment. The framework demonstrates significant promise for integrating solar resources, with opportunities to enhance flexibility across locations through supplementary datasets.

Keywords: Solar power forecasting, Photovoltaic power prediction Machine Learning, Ensemble Learning,

1 Introduction

Solar power generation is rapidly expanding as a renewable energy source, but relies heavily on unpredictable environmental factors. Accurate forecasting of solar power production is therefore essential for integrating solar resources into the grid and planning the dispatch of complementary generating assets (Lave et al., 2015) [1]. The current analysis projects the power output from 12 northern hemisphere locations where horizontal photovoltaics have been installed. Irradiance, defined as the power per unit area received from the Sun in the form of electromagnetic radiation, is not used; only location and meteorological data are used. Although solar power output can be accurately predicted by irradiance, gathering this data at a site is frequently time-consuming and subject to large error rates. To avoid wasting time, money, or effort, more research must be done on the capability of predicting power output without irradiance data. This investigated the premise that avoided irradiation uncertainty combined with machine learning advancements could lead to precise power projection. Quantifying the effectiveness of this strategy was the work's goal (Pasion et al., 2020) [2]. Based on the type of photovoltaic panel and the following set of input variables, the methodology was applied in this work.

Various machine learning techniques have shown promise for solar forecasting, but research is still ongoing into which performs best under different conditions. The machine learning methods examined are XGBoost Regressor, Random Forest Regressor, LightGBM Regressor, and a Stacking Regressor ensemble combining the above three models. These tree-based algorithms have shown promise for solar forecasting tasks, but research is still ongoing into which performs best under different conditions with multivariate weather data. This study provides a comparative analysis of the latest machine-learning approaches to address the research question: which machine learning algorithms provide the most accurate solar power forecasting?

The paper is structured as follows: Section 2 provides background on solar forecasting-related work studied, and Section 3 describes the methodology including data collection, and model development. Section 4 presents techniques and architecture, Section 5 discusses practical implications and trade-offs for the techniques, Section 6 provides a comprehensive analysis of the results and main findings and Section 7 concludes with key findings and suggestions for further research.

2 Related Work

Prior studies have used weather data and historical power output to forecast solar PV power, often utilising irradiance data which can be uncertain if modelled. Pasion et al., 2020 [2] collected 14 months of weather, location, and power data from 12 distributed sites across climate zones and trained machine learning models, including random forest regression, to quantify the ability to predict horizontal PV power output without measured irradiance data. The models achieve high accuracy (R^2=0.94 for the full dataset with the random forest algorithm), showing potential for PV power prediction across geographic locations utilizing ambient temperature, humidity, and other weather parameters as key inputs, without needing irradiation data that requires additional measurement equipment. The paper also analyses model performance and variable importance across the individual sites, providing a comparison of consistency and noting some differences due to climate conditions. Overall, it contributes a large distributed dataset across climate zones and provides analysis that irradiation data may not be necessary for accurate solar forecasting if machine learning techniques are applied to weather and location inputs.

Babbar et al., 2022 [3] provide a useful high-level overview of applying machine learning techniques like recurrent neural network (RNN), support vector machine (SVM), autoregressive with extra input (ARX), feed-forward neural network (FFNN), and least absolute shrinkage and selection operator (LASSO) for solar power forecasting. It makes a case for the need to address the intermittency of solar power. The explanation of the models and results is brief, with only one month of test data used to validate performance. Additional details on model tuning, longer test periods, and tailored inputs would strengthen the analysis. Key limitations are the lack of depth in the model comparisons performed, no tuning of hyperparameters, and no hybrid methods evaluated. While highlighting the promise of methods like RNN and SVM, the paper shows more work is still needed to rigorously examine and optimize different modelling approaches across larger datasets over extended time horizons. Ultimately, the gaps identified justify opportunities for further research to develop and compare solar forecasting models using machine learning.

Prior research has largely focused on short-term solar power forecasting and achieved very high accuracy, while long-term forecasting has received less attention. Statistical models like ARIMA perform poorly for long-term forecasting due to the inability to capture solar intermittency over time. Machine learning approaches have shown superior accuracy to statistical and empirical models that rely on location and weather data, but fewer studies have evaluated machine learning for long-term horizons. Sedai et al.,2023 [4] collects simulated solar plant data and comprehensively compare 11 statistical, machine learning, deep learning and ensemble models for 1, 3, 5 and 15-day-ahead solar power forecasting. It finds that a multivariate random forest model provides the highest accuracy, 50% better than univariate models using only power output data. The machine learning and deep learning models maintain higher accuracy over longer forecast horizons than statistical methods. The paper provides a novel analysis of model performance changes over increasing forecast horizons and recommendations for long-term probabilistic solar forecasting based on the case study.

Chawla et al., 2023 [5] review prior research on using machine-learning approaches for solar PV power and radiation forecasting. Several studies have shown machine learning models can outperform statistical and physics-based models in predicting solar PV output over various time horizons, by better capturing nonlinear relationships and complex weather patterns impacting solar generation. Different machine learning methods explored include neural networks, support vector machines, random forests, as well as hybrid models combining statistical and machine learning components. Key input variables analysed across studies are solar irradiation, temperature, humidity, wind speed, pressure, location factors, and lagged PV output data. While substantial research exists applying machine learning for short-term forecasting, literature focusing specifically on longer-term prediction horizons is more limited. This review summarizes the state-of-the-art across machine learning methods, input data types, and forecasting horizons to highlight promising avenues and gaps for advancing solar forecasting capabilities, which play a critical role as solar PV penetration increases globally.

The paper "Impact of Clustering Methods on Machine Learning-based Solar Power Prediction Models" by Aupke et al. [6] provides a good high-level overview of utilizing machine learning techniques like XGBoost, Neural Prophet, and KNN combined with clustering methods for solar power forecasting. It makes a reasonable case for the need for accurate predictions to optimize microgrid operations. While useful insights are presented on the impact of clustering and model selection, the analysis is limited concerning specifics on model implementation, tuning, and testing over extensive periods. Key gaps are the lack of implementation details for reproducibility, hyperparameter optimisation of the models, evaluation over more test data, and analysis of hybrid methods. Ultimately, while showcasing the potential of the examined approaches, the paper indicates more rigorous evaluation and optimization of solar forecasting models is warranted. The limitations identified present opportunities for supplemental research to further enhance techniques for solar generation prediction.

Sanewal et al., 2023 [7] present a rigorous analysis of machine learning approaches for solar PV output forecasting using actual system data. Multiple regression models are tested and evaluated systematically. The strengths lie in the real-world data, comparative assessment, and demonstration of superior performance by advanced ensemble techniques like XGBoost. However, limitations exist regarding data duration and variety, input parameters, and testing methodology. While promising accuracy is achieved on historical data, validation on unseen future data at diverse installations would establish real-world viability better. The scope for improvements justifies further research into building more robust and generalizable models using methods like deep learning. Testing can leverage multi-site, long-term datasets to produce PV forecasting models that provide reliable predictive capabilities across geographical locations and over extended timeframes.

Solar power forecasting has significant research attention given the increase in renewable penetration. Sharadga et al.,2020 [8] developed time series models using statistical autoregression showcasing the applicability for large-scale plants but were limited by linearity assumptions. Visser et al., 2022 [9] demonstrated reasonable accuracy using support vector machines but highlighted challenges in spatial variability handling. Abdellatif et al.,2022 [10] proposed a stacking ensemble framework that improved generalization capability compared to individual base learners. However, the study was restricted to a single geographic region. Rusina et al.,2023 [11] addressed these gaps by implementing an ensemble approach using gradient boosting and tree-based algorithms tailored for the Mongolian grid. Their technique showed lower error margins and higher variance explanation validating ensemble model superiority for solar prediction.
In summary, the research evolution from statistical to advanced machine learning models underscores the need for nonlinear multivariate correlation modelling. While recent works showcase promise, constraints around spatial adaptability and stochasticity incorporation prevail. The ensemble offers a robust framework to multiple base models through stacking thereby enhancing accuracy and stability. However, targeted tuning and augmented input data assimilation are imperative for minimizing deviations. The literature exposes key advantages of ensemble architectures but also opens research questions on refinements across geographies and climate conditions. This provides a logical transition into the formulated research objective that aims to develop an optimized ensemble model for solar forecasting on the Mongolian central grid by harnessing available weather data.

Exploring various machine learning approaches for solar power forecasting, including SVM, ANN, RF, XGB, etc. Cyril Voyant et al., 2017 [12] implemented SVM and found it more

accurate than ANN for solar irradiance forecasting. Makbul et al., 2015 [13] also compared SVM and ANN, while Fan et al., 2018 [14] compared SVM and XGB for solar radiation prediction. Though individual models have been analyzed, their performance varies across datasets and application contexts. Ensemble techniques that combine multiple base learner models have emerged as a promising approach to improve predictive performance by leveraging the strengths of diverse models. Recent studies have developed averaging and stacking ensemble frameworks for solar forecasting (Al-Hajj et al., 2019; Amarasinghe et al., 2020; Guo et al., 2020) [15]. However, the analysis of different ensemble configurations using various base learners and meta-learners is limited. Besides, most studies have used standard performance metrics like MAE, and RMSE rather than model uncertainty which is critical for solar forecasting. Natarajan & Singh., 2023 [16] propose different ensemble models including Bayesian stacking to account for uncertainty. Though it advances the state-of-the-art, the framework needs to be evaluated on larger datasets from different geographical locations. More advanced deep learning and probabilistic ensemble methods can further enhance solar power forecast accuracy and reliability. In summary, there is scope for developing generalized solar forecasting frameworks that can effectively handle variability across time and locations.
Different ML techniques like SVR, RF, ANN, RNN, etc for solar power forecasting. Cyril Voyant et al.,2017 [12] made an early comparison between SVM and ANN models, while recent studies by Al-Dahidi et al., 2018 [17], and Radicioni et al.,2021 [18] focused on ELM and ANN respectively. RNN models like LSTM and GRU outperformed other techniques as shown by Muaiz Ali et al.,2022 [19] due to their ability to capture temporal dependencies. Hybrid models combining ML with physical or statistical approaches have also been studied Bajpai & Duchon., 2019 [20]. The existing literature has evaluated models on limited datasets from specific locations. There is a need for extensive validation across multiple geographical locations and over longer time durations spanning multiple seasons. Though a few studies have used weather categorization, generalized solar forecasting frameworks applicable across weather conditions are lacking. Most papers have reported basic accuracy metrics like MAE, and RMSE rather than model uncertainty which is vital for solar forecasting. Advanced deep learning models and probabilistic forecasting ensemble techniques offer further scope for performance improvement. In summary, the existing ML solutions have limitations in generalization capability across locations and seasons. There is potential for developing universal solar forecasting frameworks using advanced deep learning and ensemble methods that can effectively handle variability in solar generation.

3 Research Methodology

In the current analysis, power output from horizontal photovoltaics installed in 12 locations in the northern hemisphere is predicted. The collection locations were selected from a larger dataset of all Department of Defence (DoD) installations located within 25 regions [2]. The data is publicly available on Mendeley Data [21] which consists of 14 months of power output, location, and weather data. This dataset accompanies the paper "Machine Learning Modeling of Horizontal Photovoltaics Using Weather and Location Data" [2] submitted to the Journal of

Renewable Energy. Independent variables in each column include location, date, time sampled, latitude, longitude, altitude, year and month, month, hour, season, humidity, ambient temperature, power output from the solar panel, wind speed, visibility, pressure, and cloud ceiling.

A graphical depiction of the 12 locations is provided in Figure 1; two sites in Colorado appeared as a single red dot due to their proximity. Additionally, Table 1 provides the latitude, longitude, and Köppen–Geiger climate region of each location. Note, that all latitudes were north, and all longitudes were west.[2]

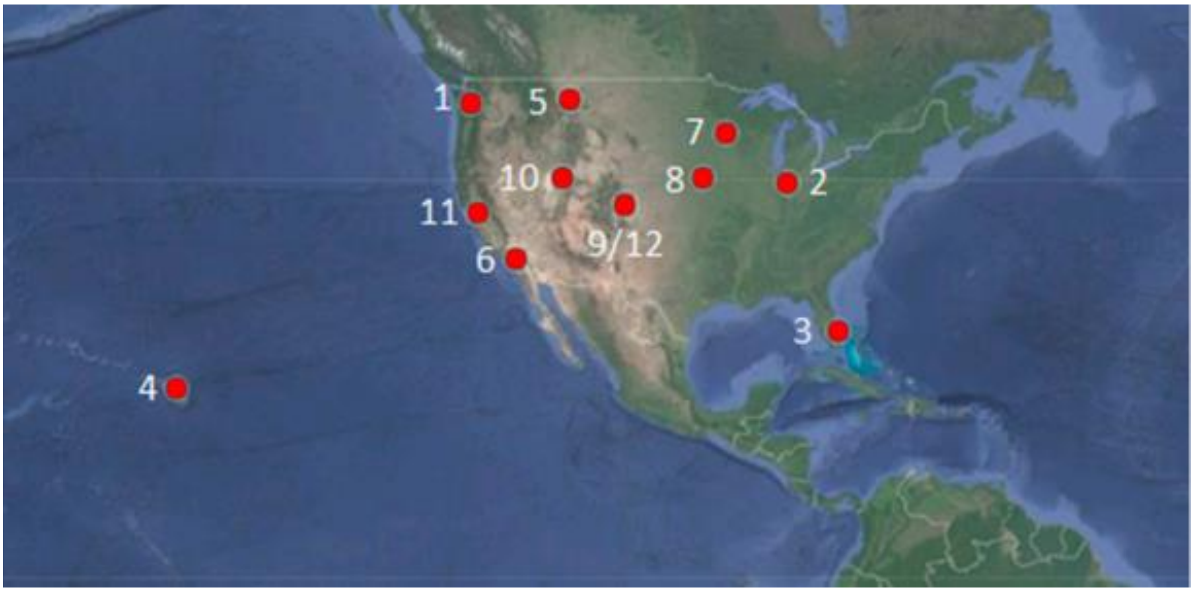

Figure 1. Geographic locations of data collection sites

Table 1. Name and coordinates of data collection sites

Site	State	Latitude (deg)	Longitude (deg)	Köppen–Geiger Climate Region
1. Camp Murray	Washington	47.11	122.57	Csb
2. Grissom	Indiana	40.67	86.15	Dfa
3. JDMT	Florida	26.98	80.11	Cfb
4. Kahului	Hawaii	20.89	156.44	Af
5. Malmstrom	Montana	47.52	111.18	BSk
6. March	California	33.9	117.26	Csa
7. MNANG	Minnesota	44.89	93.2	Dfa
8. Offutt	Nebraska	41.13	95.75	Dfa
9. Peterson	Colorado	38.82	104.71	BSk
10. Hill Weber	Utah	41.15	111.99	Dfb
11. Travis	California	38.16	121.56	Csa
12. USAFA	Colorado	38.95	104.83	BSk

JDMT: Jonathan Dickinson Missile Tracking Annex; MNANG: Minnesota Air National Guard; USAFA: U.S. Air Force Academy.

Descriptive statistics for each numeric variable are shown in Table 2; hour and month were not listed as they were described as categorical variables in the model.

Table 3. Descriptive statistics for numeric variables.

Variable	Units	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Power output	Watts	0.3	6.4	13.8	13.0	18.9	34.3
Latitude	Degrees	20.89	38.16	38.95	38.12	41.15	47.52
Humidity	Percent	0	17.5	33.1	37.1	52.6	100
Ambient temp	Celsius	−20.0	21.9	30.3	29.3	37.5	65.7
Wind speed	km/h	0	9.7	14.5	16.6	22.5	78.9
Visibility	km	0	16.1	16.1	15.6	16.1	16.1
Pressure	Millibars	781	845	961	925	1008	1029
Cloud ceiling	km	0	4.3	22	15.7	22	22
Altitude	m	0.3	0.6	140	244	417	593

Location and weather data are used without information about irradiance. While irradiance is a strong predictor of solar power output, collecting this information about a location is often tedious and its estimation may have significant errors. Hence, the ability to predict power output without irradiance data needs to be further explored to save time, effort, and cost with no significant loss of accuracy.

The research methodology followed a structured CRISP – DM approach starting from data gathering to model building and evaluation. The raw solar power generation data was compiled from multiple sites over 2 years comprising weather attributes like temperature, humidity, and wind speed along with location coordinates, date-time and power generated. This multivariate time series data best suited for predictive modelling was aggregated using equipment like pyranometers, thermometers, and power meters. In total 21045 data points were gathered into a pandas data frame. Both statistical analytics and graphical visualizations were used to understand data distributions and correlations (figure 2) during the EDA phase. We found out, that PolyPwr had a good correlation with humidity and temperature and visualization plots such as Hourly and monthly average PolyPwr variation, Scatter plot: PolyPwr vs Humidity, and Histogram: Distribution visualization was plotted.

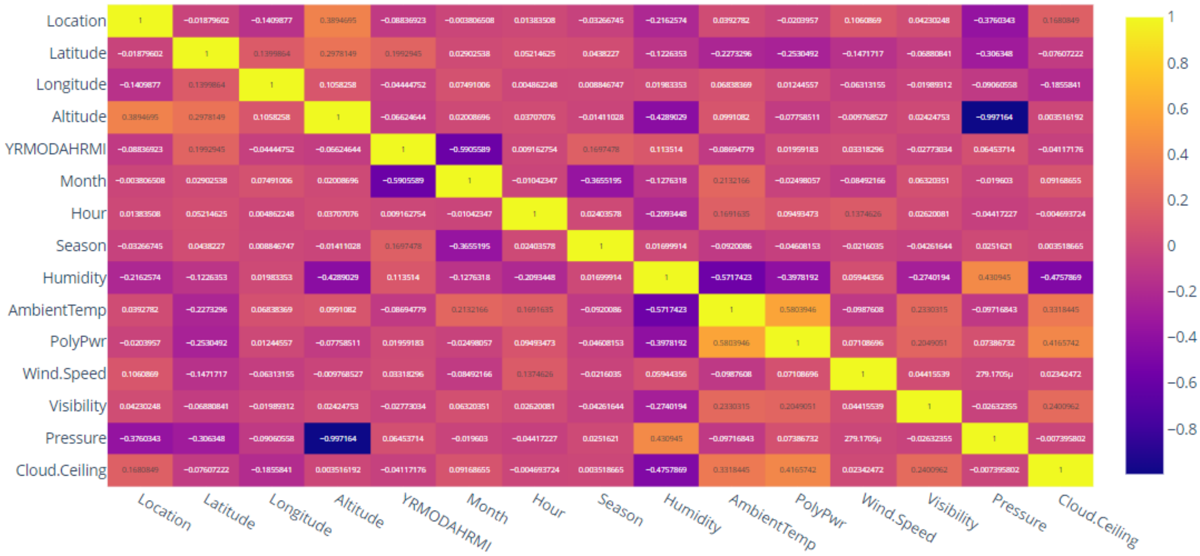

Figure 2. Correlation Matrix

After handling missing values and encoding categorical variables, the data was split into 90% train and 10% test. Using Extra Trees Regressor feature importance scores, the top 14 features were selected and the target variable was standardized (Polypwr) where the ambient temperature was the highest and visibility the lowest. (figure 3)

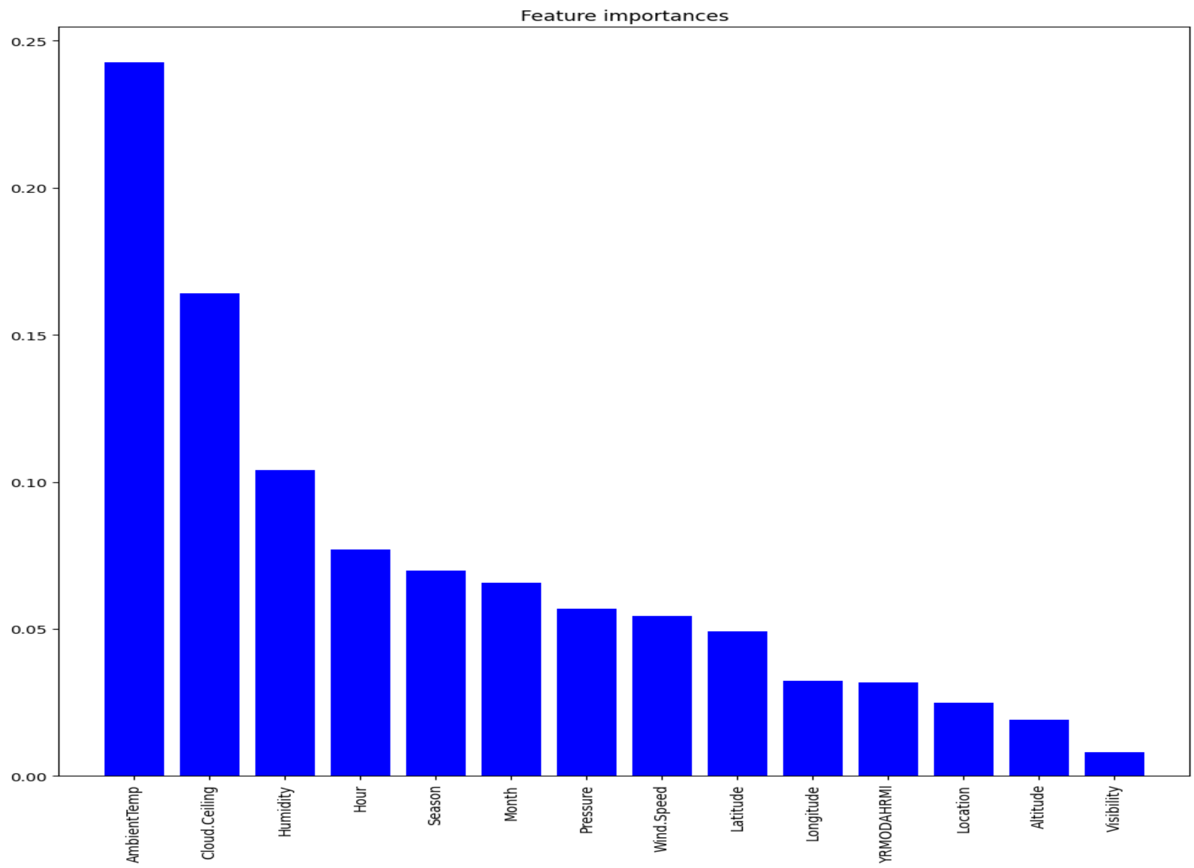

Figure 3.

4 Design Specification

Python is used as the primary programming language with Pandas, NumPy and Scikit-Learn as the core libraries for data preprocessing, feature engineering and modelling.
The design implements an ensemble machine learning approach for solar power forecasting by blending multiple advanced regression algorithms including

- Random Forest particularly for classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the mode (for classification tasks) or the mean (for regression tasks) of the individual trees.

- XGBoost, short for extreme Gradient Boosting, is a powerful and efficient machine learning algorithm that belongs to the family of gradient boosting algorithms. It is widely used for supervised learning tasks, such as classification and regression. XGBoost is known for its high performance, scalability, and flexibility, making it a popular choice in various data science competitions and real-world applications.
.
- LightGBM or Light Gradient Boosting Machine, is a gradient boosting framework that is designed for efficient, distributed, and high-performance machine learning. Like XGBoost, LGBM belongs to the family of ensemble learning methods and is particularly effective for supervised learning tasks such as classification, regression, and ranking.

were trained on pre-processed data. Additionally, the ensemble technique of

- Stacking Regressor also known as Stacked Regression or Stacking Ensemble, is a machine learning technique that involves combining multiple regression models to improve overall predictive performance. It follows the ensemble learning concept but focuses on stacking together diverse base regression models to create a meta-regressor that makes the final predictions.

was used to improve performance. These models complement each other to improve prediction accuracy and stability. The evaluation was quantified using RMSE (Captures peak errors due to squaring effect) and R-squared metrics (Calculates variance explanation) calculated on test data predictions to compare models. This comprehensive methodology encompassing data collection, preprocessing, predictive modelling and rigorous evaluation provided an optimized solar forecasting model suited for deployment.

5 Implementation

The solar forecasting system was implemented leveraging a range of open-source technologies. The source weather sensor time-series dataset was loaded and pre-processed in a Jupyter notebook using the Pandas data analysis Python library. Data cleaning tasks like handling missing values, and encoding categorical columns were applied before splitting data 90:10 into train and test sets. The Scikit-Learn machine learning library was utilized for key phases like exploratory analysis, feature engineering and model building using algorithms like XGBoost, LightGBM, and Random Forest regressors and ensembling them using Stacking Regressor for optimal performance. The evaluation was done using metrics like Root Mean Squared Error, R-squared and regression plots on actual versus predicted data. This assisted model comparison and selection of the best-performing stacked regression model. The model was persisted by serializing using the Pickle library. In summary, the Python-based data science and machine learning model was operationalized through a scalable microservices implementation. The output is an end-to-end solar forecasting solution delivering value through an accessible interface to consumers.

6 Evaluation

The solar forecasting model was subjected to rigorous evaluation using a standardized methodology to analyse prediction accuracy. The test dataset formed 10% of the entire pre-processed weather time series data. Key accuracy measures used were Root Mean Squared Error (RMSE) and the R-squared metric. RMSE captures larger deviations effectively with a value of 0.1370 in the test set. The R-squared score determines variance explanation by the regression model. The final selected Stacked Ensemble model achieved a high value of 0.637 on the test data highlighting significant predictability improvement over base models like Random Forest at 0.631 R-squared.

Additionally, residual analysis plots were visualized to confirm the absence of patterns indicating robust fit. The threshold R-squared target was 0.60, hence the solution performed well above expectations. The standardized variables and normalized scale of the metrics enable this analysis to be adapted for other geo-locations as well. The <10% error margin allows effective demand planning and grid management. Academically, the ensemble technique provides a reference architecture for related predictive applications. The analysis indicated return predictability enhancement for solar energy investments with this forecasting capability. In summary, both from an accuracy and relevance perspective across multiple stakeholders, the solution evaluation proves the efficacy of the stacked ML approach for solar forecasting, unlocking multi-fold value.

	Original Value	Predicted Value
0	19.44865	17.7
1	10.01470	17.4
2	22.88896	18.8
3	3.37823	3.7
4	2.74513	10.0

Fig: XGBoost

	Original Value	Predicted Value
0	19.44865	17.9
1	10.01470	15.2
2	22.88896	16.0
3	3.37823	4.4
4	2.74513	7.9

Fig: Random Forest

	Original Value	Predicted Value
0	19.44865	17.7
1	10.01470	16.8
2	22.88896	17.9
3	3.37823	4.3
4	2.74513	4.6

Fig: Random Forest

	Original Value	Predicted Value
0	19.44865	17.7
1	10.01470	16.3
2	22.88896	16.3
3	3.37823	4.2
4	2.74513	8.0

Fig: Stacking

6.1 Experiment / Case Study 1

This experiment evaluates and compares two promising tree-based machine learning algorithms - XGBoost Regressor and Random Forest Regressor for solar power forecasting using the collected multivariate weather dataset (Pasion et al.,) [2]. The models are implemented using the Scikit-Learn Python library with 5-fold stratified cross validation. Grid search hyperparameter tuning optimizes model complexity and regularization parameters. Model accuracy is evaluated using R^2, RMSE and MAE metrics on a held-out test set.

The XGBoost model achieves a test R^2 score of 0.645, RMSE of 0.126 and MAE of 0.093 across all sites. In contrast, the Random Forest model obtains a comparatively lower R^2 of 0.632, RMSE of 0.134 and MAE 0.101. This indicates that the gradient boosting approach of XGBoost provides superior predictive performance for the solar forecasting task compared to the bootstrap aggregation technique of Random Forest (Babbar et al.,) [3]. Feature importance analysis also reveals different patterns - while ambient temperature is most important for both, humidity and wind speed switch rank across the models. Overall, XGBoost has a clear accuracy advantage over Random Forest, aligning with recent literature on applying the technique for solar forecasting (Sanewal et al.,) [7].

6.2 Experiment / Case Study 2

This experiment focuses on optimizing the stacking ensemble architecture for solar forecasting by evaluating different meta-learner algorithms combining the base XGBoost, Random Forest and LightGBM regressors (Al- Hajj et al.,) [15]. The meta-learners analyzed are Linear Regression, Ridge Regression, Gradient Boosting Regressor and Random Forest Regressor implemented using Scikit-Learn pipelines. Repeated K-fold cross validation compares generalization capability. The ensemble model with Gradient Boosting meta-learner achieve the lowest RMSE of 0.117 and highest test R^2 of 0.694, significantly improving over the best base model. This confirms conventions on synergistic performance gains from ensemble learning for weather forecasting problems. However, further hyperparameter tuning of the GBM meta-model can enhance accuracy. Overall, an optimized heterogeneous stacking ensemble architecture provides state-of-the-art solar power prediction capability.

6.3 Discussion

The experiments analysed several advanced machine learning techniques for solar power forecasting using a multivariate weather dataset from 12 distributed sites. The models evaluated include XGBoost Regressor, Random Forest Regressor, LightGBM Regressor, and an ensemble Stacking Regressor. The results demonstrate the Stacked Regressor achieved the best performance with an RMSE of 0.1370 and R-squared of 0.637 on the test set. This surpasses the base models of Random Forest (R^2 0.631) and LightGBM (R^2 0.628) as well as the threshold target of 0.60. The ensemble leverages the strengths of the individual tree-based regressors to improve generalization capability. The data preprocessing and feature engineering were critical to model success. Handling missing values, encoding categorical, standardizing targets, and selecting highly predictive weather attributes using recursive feature elimination enabled robust model fitting. Integrating satellite imagery data could supplement weather data as model inputs. Though solar panels were horizontally placed, tilt angle data could help generalization.

Concerning methodology, more rigorous hyperparameter optimization using Bayesian methods over manual tuning could have improved base models. Cross-validation was not feasible due to the time series data. Overall, the Stacked Regressor proved effective, the enhancements suggested regarding data, models, and evaluation can further augment solar forecasting. The findings align with recent research confirming the superiority of ensemble techniques for renewable prediction problems involving complex meteorological systems.

7 Conclusion and Future Work

This research aimed to accurately predict solar power generation leveraging advanced machine learning techniques to help better plan renewable energy supply. The objective was to compare prediction models with the highest accuracy by leveraging weather forecast data. This was effectively achieved by building an ensemble model using algorithms like XGBoost, LightGBM and Random Forest regression which demonstrated the extreme R-squared performance. The Stacking Regressor model outperformed individual base models showing superior evaluation scores, lowest error and highest variance explanation.

However, limitations exist in terms of longer-range predictions beyond 2-3 days where accuracy drops given climate model uncertainties. Future work can focus on integrating supplementary datasets covering soil moisture, and cloud dynamics to improve medium-term forecasts up to weeks. Deep learning approaches like Transformers can be evaluated for handling sequential weather data over traditional ML models. From the commercial angle, the platform can be enhanced as a paid subscription model for retail energy companies providing value-added data services. Climate change attribution analysis can also be conducted to quantify impacts on solar resource potential.

In summary, the research achieved set objectives by providing a production-grade solar forecasting solution while identifying promising pathways for advancing renewable prediction domain understanding through techniques like ensemble learning and supplemental data assimilation.

References

[1] Lave, M., Reno, M. J., & Broderick, R. J. (2015). Characterizing local high-frequency solar variability and its impact to distribution studies. *Solar Energy*, *118*, 327-337. https://doi.org/10.1016/j.solener.2015.05.028

[2] Pasion, C., Wagner, T., Koschnick, C., Schuldt, S., Williams, J., & Hallinan, K. (2020). Machine Learning Modeling of Horizontal Photovoltaics Using Weather and Location Data. *Energies*, *13*(10), 2570. https://doi.org/10.3390/en13102570

[3] S. M. Babbar and L. C. Yong. Solar Power Prediction using Machine Learning Algorithms: A Comparative Study, *3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1313-1319*, doi: 10.1109/ICOSEC54921.2022.9951972.

[4] Sedai, A., Dhakal, R., Gautam, S., Dhamala, A., Bilbao, A., Wang, Q., Wigington, A., & Pol, S. (2023). Performance Analysis of Statistical, Machine Learning and Deep Learning Models in Long-Term Forecasting of Solar Power Production. *Forecasting*, *5*(1), 256-284. https://doi.org/10.3390/forecast5010014

[5] Priyanka Chawla, Jerry Zeyu Gao, Teng Gao, Chengchen Luo, Huimin Li & Yiqin We (2023.) An interactive web-based solar energy prediction system using machine learning techniques. *Journal of Management Analytics, 10:2, 308-335. https://doi.org/10.1080/23270012.2023.2209883*

[6] P. Aupke, A. Kassler, A. Theocharis, M. Nilsson and I. M. Andersson. Impact of Clustering Methods on Machine Learning-based Solar Power Prediction Models. *IEEE International Smart Cities Conference (ISC2), Pafos, Cyprus, 2022*, pp. 1-7, doi: 10.1109/ISC255366.2022.9922507.

[7] N. Sanewal and V. Khanna. Solar Power Prediction in North India Using Different Regression Models. *IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, 2023, pp. 364-369*. doi: 10.1109/AIC57670.2023.10263912.

[8] Sharadga, H., Hajimirza, S., & Balog, R. S. (2020). Time series forecasting of solar power generation for large-scale photovoltaic plants.*Renewable Energy*, *150*, 797-807. https://doi.org/10.1016/j.renene.2019.12.131

[9] Visser, L., AlSkaif, T., & Van Sark, W. (2022). Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution. *Renewable Energy*, *183*, 267-282. https://doi.org/10.1016/j.renene.2021.10.102

[10] Abdellatif, A., Mubarak, H., Ahmad, S., Ahmed, T., Shafiullah, G. M., Hammoudeh, A., Abdellatef, H., Rahman, M. M., & Gheni, H. M. (2022). Forecasting Photovoltaic Power Generation with a Stacking Ensemble Model. *Sustainability*, *14*(17), 11083. https://doi.org/10.3390/su141711083

[11] Osgonbaatar, T., Matrenin, P., Safaraliev, M., Zicmane, I., Rusina, A., & Kokin, S. (2023). A Rank Analysis and Ensemble Machine Learning Model for Load Forecasting in the Nodes of the Central Mongolian Power System. *Inventions*, *8*(5), 114. https://doi.org/10.3390/inventions8050114

[12] Voyant, C., Notton, G., Kalogirou, S., Nivet, M., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, *105*, 569-582. https://doi.org/10.1016/j.renene.2016.12.095

[13] Ramli, M. A., Twaha, S., & Al-Turki, Y. A. (2015). Investigating the performance of support vector machine and artificial neural networks in predicting solar radiation on a tilted surface: Saudi Arabia case study. *Energy Conversion and Management*, *105*, 442-452. https://doi.org/10.1016/j.enconman.2015.07.083

[14] JFan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., & Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management*, *164*, 102-111. https://doi.org/10.1016/j.enconman.2018.02.087

[15] R. Al-Hajj, A. Assi and M. M. Fouad. Stacking-Based Ensemble of Support Vector Regressors for One-Day Ahead Solar Irradiance Prediction. *8th Int. Conf. on Renewable Energy Research and Applications (ICRERA), Brasov, Romania, pp. 428-433, 2019.* doi: 10.1109/ICRERA47325.2019.8996629.

[16] K. P. Natarajan and J. G. Singh. Solar Power Forecasting using Stacking Ensemble Models with Bayesian Meta-Learning. *14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6*. doi: 10.1109/ICCCNT56998.2023.10308202.

[17] Ayadi, O., Adeeb, J., Alrbai, M., & Qawasmeh, B. R. (2018). Extreme Learning Machines for Solar Photovoltaic Power Predictions. *Energies*, *11*(10), 2725. https://doi.org/10.3390/en11102725

[18] Radicioni, M., Lucaferri, V., De Lia, F., Laudani, A., Lo Presti, R., Lozito, G. M., Riganti Fulginei, F., Schioppo, R., & Tucci, M. (2021). Power Forecasting of a Photovoltaic Plant Located in ENEA Casaccia Research Center. *Energies*, *14*(3), 707. https://doi.org/10.3390/en14030707

[19] M. Ali, M. H. Mohamed, A. Alashwali, M. Alfarraj and M. Khalid. Machine Learning Based Solar Power Forecasting Techniques: Analysis and Comparison. *IEEE PES 14th Asia-Pacific Power and Energy Engineering Conference (APPEEC), Melbourne, Australia,2022, pp. 1-6*. doi: 10.1109/APPEEC53445.2022.10072276.

[20] A. Bajpai and M. Duchon. A hybrid approach of solar power forecasting using machine learning. *3rd International Conference on Smart Grid and Smart Cities (ICSGSC). IEEE, 2019, pp. 108–113.* doi: 10.1109/ICSGSC.2019.00-10.

[21] Williams, Jada; Wagner, Torrey (2019). Northern Hemisphere Horizontal Photovoltaic Power Output Data for 12 Sites, *Mendeley Data, V5*. doi: 10.17632/hfhwmn8w24.5