

Article

Not peer-reviewed version

---

# Combined K-means Clustering with Neural Networks Methods for PV Short-Term Generation Load Forecasting in Electric Utilities

---

[Alex Sleiman](#)<sup>\*</sup> and [Wencong Su](#)

Posted Date: 15 November 2023

doi: 10.20944/preprints202311.0977.v1

Keywords: machine learning; neural networks; PV power forecasting; smart meters; solar energy



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# Combined K-Means Clustering with Neural Networks Methods for PV Short-Term Generation Load Forecasting in Electric Utilities

Alex Sleiman <sup>1,2,\*</sup> and Wencong Su <sup>2</sup>

<sup>1</sup> DTE Electric, Detroit, MI 48226, USA

<sup>2</sup> Department of Electrical and Computer Engineering, University of Michigan-Dearborn, Dearborn, MI 48128, USA; wencong@umich.edu

\* Correspondence: ahsleima@umich.edu

**Abstract:** The power system has rapidly grown and expanded over the past decades and has been experiencing major changes and challenges. The increase in energy demand and the modern advancements in the smart grid, such as solar and wind energies and electric vehicles, have led to complexity and complications for utilities. A further layer of complexity and difficulty was added by the rapid expansion of behind-the-meter (BTM) photovoltaic (PV) systems with various designs and characteristic features. The rapid increase and the invisible solar power (BTM) have led to fluctuations in power grid stability and reliability and to inefficiency. Accurate forecasting of load generation will help to assure optimal planning, minimize the negative effects of the PV systems, and minimize the operational and maintenance costs. The authors propose a solution that uses combinations of K-means clustering with neural network machine learning models, AMI real-world PV load generation, and weather data to forecast the generation load at customer locations to achieve a 2.49% error between actual and predicted generation load.

**Keywords:** machine learning; neural networks; PV power forecasting; smart meters; solar energy

## 1. Introduction

Electricity generation systems have traditionally been dominated by fossil fuel-based generators. Yet, because of these generators' detrimental effects on the environment, the power industry is currently concentrating on green energy-based alternative generation technologies [1]. Solar energy photovoltaics (PVs) is one of the most desirable green energies among the many renewable energy resources since it has environmental and economic benefits [2]. In addition, the U.S. Department of Energy Solar Energy Technologies Office and the National Renewable Energy Laboratory (NREL) developed a solar future study, the findings of which were that solar will grow from 3% of the U.S. electricity supply today to 40% by 2035, 95% decarbonized, and 45% by 2050 to achieve 100% decarbonization [3]. However, the integration of solar PVs has added significant issues to the power grid, such as system stability, reliability, electric power balance, reactive power compensation, and frequency response. Solar PV generation forecasting has emerged as an effective solution to address these problems [4,5].

This manuscript is designed for the electric utility industry. It explores the importance of precise solar forecasting in various aspects of photovoltaic (PV) system management and operation. The accurate prediction of solar energy output offers numerous benefits, including the reduction of auxiliary device maintenance costs, increased PV system penetration, mitigation of the impact of uncertain solar PV output, and enhanced overall system stability. By utilizing advanced approaches such as neural network machine learning models, AMI real-world PV generation data, and weather data, the proposed methodologies for PV solar load forecasting provide additional industry-specific advantages. These include enhanced accuracy in load estimation, facilitating optimal planning and resource allocation. Moreover, the improved forecasting accuracy helps minimize operational and maintenance costs, ensuring efficient utilization of PV systems. Furthermore, it enables better

integration of renewable energy sources like solar PV, optimizing their contribution to the overall energy mix and promoting a more sustainable energy landscape. With the application of these approaches, the industry can achieve better operational efficiency, cost savings, and a reliable power grid while harnessing the full potential of solar energy resources. The key areas of focus in this study are the development of forecasting methodologies for solar PV power generation. These methodologies aim to optimize the utilization of solar energy by effectively predicting the amount of power that can be generated from PV systems. The forecasting models employed in this research consider several factors such as geographical location, climatic variability, input parameter selection, and training procedures.

To achieve accurate solar forecasting, machine learning techniques are analyzed and compared in this manuscript. Machine learning has proven to be an effective tool in capturing complex patterns and relationships within solar data, enabling more precise predictions. By evaluating and comparing different machine learning algorithms, this study aims to identify the most suitable approach for solar PV generation load forecasting.

To ensure the effective integration of solar power into the energy grid, fostering grid stability, advancing toward a more sustainable future and optimal planning, the contributions of this paper are as follows:

- (1) The model is specifically tailored for electric utilities. It is intended to be integrated into electric utility systems that utilize automated meter infrastructure (AMI) data for their operational processes.
- (2) We utilized real electric utility AMI data collected from customer locations in the proposed models.
- (3) We reduced the error between actual and forecasted load generation from 14.3% to 2.49% in all weather conditions and achieved excellent results in comparison with the best models.
- (4) We compared the model's performance with multiple widely used algorithms in the context of photovoltaic power forecasting.

The remainder of this manuscript is organized as follows: In Section 2, we highlight related work in solar generation forecasting techniques. In Section 3, we present our methodology, which consists of data collection and forecast models. In Section 4, we give the results for each model and evaluate the performance. In Section 5, conclusions and future work are discussed. References used in the article are in the last section.

## 2. Related Work

Energy forecasting is a well-known and long-standing concern in power systems. Nevertheless, with advances in artificial intelligence and machine learning over the past ten years, as well as the expansion of distributed energy resources, load forecasting has emerged as a key issue. There are different methodologies for energy forecasting. They can be classified into time-series models, regression models, and artificial neural network models [6].

Time-series models like autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) have lower data requirements and less computation cost [7]. However, the main challenges of these models are high complexity of time series data, low accuracy, and poor generalization ability of the prediction model [8]. These models were proposed in [9].

Regression models can be developed for different time ranges (short, medium, and long term) considering different influential environmental and temporal parameters [10]. Linear regression (LR) models are widely used in energy forecasting since they are simple and effective [11]. Another regression technique is support vector regression (SVR); it was developed to provide a relationship between two nonlinear variables. At times even SVR also fails to provide better results while considering the weather parameters [12]. In [13] and [14] the regression models including SVR were discussed and proposed. Artificial neural network (ANN) models are most used when dealing with complex problems or nonlinear patterns [15]; thus, these models were explored for solar power forecasting and are applied in [16–18].

Deep learning or deep neural network (DNN) has emerged in recent years; it is a development of neural networks with improved capacity to solve complex problems. A deep learning framework to predict PV power forecasting was discussed in [19,20]. Long short-term memory (LSTM) is a type

of deep learning network that has been widely utilized in time series forecasting and has shown powerful results [21]; this type of deep learning has been proven and demonstrated as highly effective models for several challenging learning tasks [22].

While many approaches have been proposed for load forecasting, there are still areas that have not been fully explored. Some potential areas for further development in load forecasting include incorporating more diverse data sources that can be leveraged, such as social media or building occupancy data, to improve forecasting accuracy. Load forecasting models assume that the statistical properties of the load data remain constant over time, but load patterns can change due to changes in consumer behavior, technological advances, or other factors. Developing methods to account for non-stationarity in load data could improve forecasting accuracy. Uncertainty is inherent in load forecasting, and incorporating methods to quantify and address this uncertainty could improve decision-making. Although load forecasting has traditionally been used for energy management, it could potentially be useful in other areas such as traffic management or supply chain optimization.

Table 1 compares different methods used in top-cited journals for short-term solar generation load forecasting. The methods are evaluated on four criteria: Whether the models used smart meter data, whether the research data were obtained from multiple locations, the forecast horizon, and the limitations of each article.

Our model uses AMI smart meter data, which is actual real-world data from customer locations, unlike the others, which used sample and device logger data. AMI smart meters measure and communicate actual load data at a customer-location level, which is more reliable and accurate. Also, we collected data from multiple geographical locations for an entire year, while some other researchers have their method applied to only one location, as shown in Table 1.

The proposed model forecast horizon is 168 hours ahead or 7 days, which is longer than that of other models. It enables electric utilities to achieve better planning and resource allocation and helps operators in balancing the grid with other energy sources. The models in Table 1 present various limitations that impact the effectiveness of their forecasting methodologies, which is essential for accurate solar generation load predictions. The model in [23] fails to consider the crucial aspect of weather data, which is essential for accurate solar load predictions. In contrast, the model in [24] generates weather forecasts, but this approach introduces uncertainties and errors in the forecasting process. The model in [25] highlights a significant challenge of dealing with a large number of missing data, leading to limited evaluation of predictions during cloudy and rainy days. Additionally, the reliance of the model in [26] on expensive and complex numerical weather prediction (NWP) data raises concerns about the practicality and cost-effectiveness of the approach.

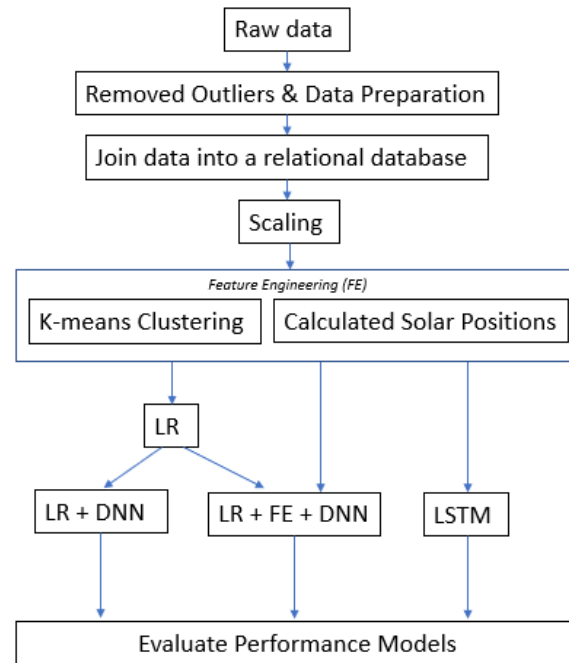
The prediction performance in model [28] being worse on typhoon days calls attention to the need for more comprehensive data collection and consideration of extreme weather events. In model [29], the use of automatic data loggers is constrained by limited memory capacity and a lack of real-time data, potentially limiting the timeliness and accuracy of predictions. Moreover, the model in [30] has insufficient training data samples and high computational processing, which negatively impacts the forecasting model's performance and practicality. In the model in [31], the use of weather satellite images introduces complexity and expense, making it challenging for widespread implementation. Finally, the model in [32] has an inadequacy of training data samples, which directly affects the accuracy of forecasts, stressing the importance of sufficient data for robust predictions. In summary, each article's limitations underscore the necessity of addressing data availability, data quality, and methodological choices to improve the reliability and applicability of solar generation load forecasting models. Our model uses the actual real-world data that utilities utilize for their day-to-day operations.

**Table 1.** Comparison of short-term generation load methodologies in top-cited journals.

Ref #	Forecasting model	Smart meter data	Multiple locations	Forecast horizon	Limitations
[23]	Probabilistic	no	yes	1 h ahead	Did not use weather data
[24]	K means Clustering with LSTM	no	no	12 h ahead	Created weather forecast
[25]	ARX	yes	yes	2 h ahead	Per the author: large number of missing data, not enough data to evaluate cloudy and rainy days.
[26]	Gradient Boosting trees	no	yes	72 h ahead	Uses NWP data, which is complicated and highly expensive [27]
[28]	SOM and Fuzzy	yes	no	24 h ahead	Per the author: forecasts worse on typhoon days, need more data collection
[29]	SVR	no	no	24 h ahead	Used automatic data logger, which are limited memory capacity and lack of real time data
[30]	Gaussian Process Regression	no	no	< 1 h	Per the author: only 33 days used for training which affected the performance, high computational process
[31]	Autoregression	no	yes	6 h ahead	Per the author: used weather satellite images which are complex and expensive
[32]	NN Ensemble	no	no	24 h ahead	Per the author: not sufficient training data samples which affected the forecast performance
<b>This article</b>	<b>K means clustering with LR DNN and LSTM</b>	<b>yes</b>	<b>yes</b>	<b>168 h ahead</b>	<b>Uses actual electric utility data</b>

### 3. Methodology

The purpose of this study is to propose a precise approach for solar PV power forecasting based on AMI load data and weather variables such as temperature, dew point, cloud coverage, and barometric pressure. Accurate forecasting methods are needed because these uncertainties affect PV power generation, which ultimately affects system stability. Figure 1 shows the flowchart diagram for the overall methodology.



**Figure 1.** Flowchart diagram of the methodology.

### 3.1. Data Collection

The historical hourly actual generation load data were obtained from an electric utility in the U.S. Midwest region. The utility provided local weather information measured from weather stations at 12 different local airports. The data covered a period of 24 months, from January 2020 to December 2021. The weather parameters contained in the dataset are temperature, barometric pressure, dew point, and cloud cover measured in oktas. In addition, the following calculated solar hourly sun positions in the sky were computed: the elevation angle in radians, the declination angle in radians, the azimuth angle in radians, and a Boolean variable denoting if the sun was above the horizon. Also, outliers were removed from the load data by month and by season using a box plots method. The data are not available to the public due to information security for the utility's customers; however, a sample of scaled data is shown in Figure 2.

### 3.2. Customer Grouping

To improve the overall accuracy of the forecasting, a K-means clustering algorithm has been used to group the customers into different clusters based on their historical solar generation and weather factors.

The algorithm used 2020 and 2021 historical time series data to group the customers. The optimum number of clusters into which the customers can be grouped has been determined by the elbow method. Four clusters have been determined by the elbow method as the optimum number to group the customers into. The clusters along with the calculated solar positions are called feature engineering (FE); they have been used as features in the models to forecast load.

### 3.3. Forecasting Methods

The weather parameters, calculated solar positions, and K-means clustering assignments were used as input features to train several supervised learning models to predict solar generation. Since there was data available from only 2 years, each model was trained using data from 2021 and was tested using 2020 data.



Date and Time	Meter Number	Generation (kw)	Above Horizon	Declination (rad)	Azimuth (rad)	Elevation (rad)	Temp	BP	DEW	Cloud (oktas)
7/3/2021 1:00	123456	0	0	0.9903	0.0010	0.3241	0.6838	0.6016	0.7480	0
7/3/2021 2:00	123456	0	0	0.9903	0.0456	0.3426	0.6838	0.6056	0.7398	0
7/3/2021 3:00	123456	0	0	0.9903	0.0865	0.3794	0.6618	0.6046	0.7236	0
7/3/2021 4:00	123456	0	0	0.9903	0.1230	0.4307	0.6618	0.6006	0.7317	0
7/3/2021 5:00	123456	0	0	0.9903	0.1553	0.4929	0.6544	0.5996	0.7154	0
7/3/2021 6:00	123456	0	1	0.9903	0.1845	0.5626	0.6397	0.5996	0.7073	0
7/3/2021 7:00	123456	0.0111	1	0.9903	0.2122	0.6375	0.6471	0.6096	0.7154	0
7/3/2021 8:00	123456	0.0733	1	0.9903	0.2400	0.7153	0.7059	0.6076	0.7642	0
7/3/2021 9:00	123456	0.1637	1	0.9903	0.2706	0.7941	0.7500	0.6076	0.7561	0.4375
7/3/2021 10:00	123456	0.1878	1	0.9903	0.3083	0.8713	0.7500	0.6056	0.7480	1
7/3/2021 11:00	123456	0.1926	1	0.9903	0.3627	0.9418	0.7574	0.6026	0.7561	1
7/3/2021 12:00	123456	0.2775	1	0.9903	0.4521	0.9923	0.7721	0.5976	0.7642	1
7/3/2021 13:00	123456	0.6966	1	0.9903	0.5757	0.9977	0.7941	0.5856	0.7724	0.75
7/3/2021 14:00	123456	0.7456	1	0.9903	0.6737	0.9538	0.8235	0.5756	0.7967	0.75
7/3/2021 15:00	123456	0.8022	1	0.9903	0.7333	0.8858	0.8529	0.5646	0.8049	0
7/3/2021 16:00	123456	0.7014	1	0.9903	0.7733	0.8094	0.8676	0.5596	0.8130	0.4375
7/3/2021 17:00	123456	0.6565	1	0.9903	0.8048	0.7306	0.8750	0.5546	0.8374	0
7/3/2021 18:00	123456	0.5017	1	0.9903	0.8329	0.6524	0.8824	0.5516	0.8374	0.4375
7/3/2021 19:00	123456	0.3238	1	0.9903	0.8605	0.5769	0.8824	0.5516	0.8455	0.75
7/3/2021 20:00	123456	0.1278	1	0.9903	0.8893	0.5059	0.8603	0.5526	0.8455	0
7/3/2021 21:00	123456	0.0247	1	0.9903	0.9209	0.4421	0.8235	0.5516	0.8455	0
7/3/2021 22:00	123456	0.0002	0	0.9903	0.9565	0.3884	0.8015	0.5536	0.8455	0.75
7/3/2021 23:00	123456	0.0000	0	0.9903	0.9965	0.3485	0.7868	0.5546	0.8374	0.4375

**Figure 2.** One day sample scaled data used for a given meter.

### 3.3.1. Linear Regression (LR)

A simple linear regression model was created to use an input for the deep neural networks model. Linear regression is a widely used method for load forecasting; it is a simple and easy-to-use method. Linear regression assumes a linear relationship between the dependent variable (load) and the independent variable (such as time or weather data), which is appropriate for load forecasting as load tends to follow a consistent trend over time. In addition, linear regression is efficient and able to handle large datasets, making it a practical option for load forecasting.

### 3.3.2. Deep Neural Networks (DNN)

Next, two fully connected feedforward neural networks were used to predict solar generation. Neural networks are preferred for short-term load forecasting due to their ability to model nonlinear relationships between variables, adapt to changes in load patterns, handle noisy and irregular data patterns, scale to large and complex datasets, and provide highly accurate forecasts.

The first model used only the output of the linear regression model as an input to predict solar load.

The second model also used the output of the linear regression model along with feature engineering FE to predict solar load. Each model used root mean squared error and mean absolute error and full batch gradient descent to optimize the weights.

Each model used two hidden layers with the ReLU function as the activation function between layers. A grid search was used to search for the optimum layer sizes. The model was trained 5 times with different random starting states for each layer size option to ensure consistency of the results.

### 3.3.3. Long Short-Term Memory (LSTM)

Finally, an LSTM model was created to predict solar generation. LSTM (long short-term memory) is a type of neural network that is well-suited for handling sequence data and has been used for load forecasting due to its ability to capture long-term dependencies and temporal patterns in the

data. LSTMs have a memory component that allows them to retain and use information from past observations in the sequence, which is useful for forecasting load, as load can be influenced by historical trends. Like other neural networks, LSTMs can model nonlinear relationships between variables, which is important in load forecasting as load can be influenced by various nonlinear factors such as weather conditions.

Rather than using features, sequential models like LSTM look at the previous solar generation values to predict the next most likely value in the future. The model was run in sequence to vector mode, where a sequence of the previous nine values was used to predict generation for one timestep into the future.

Root mean squared error and mean absolute error were used to evaluate the performance as the model used a single LSTM layer. The model was trained using full batch gradient descent. A grid search methodology as described above was used to choose the optimum LSTM layer and window size.

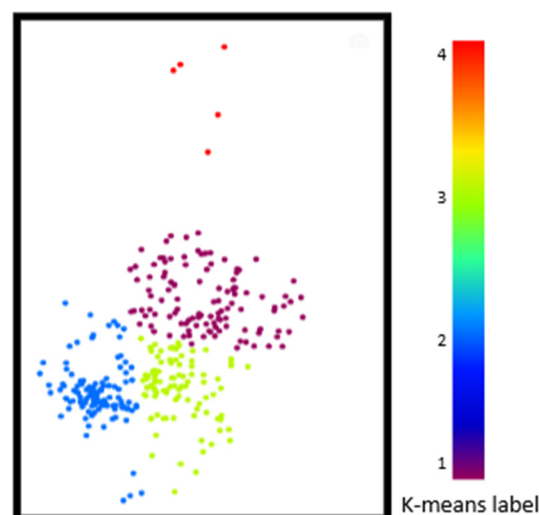
### 3.3.4. Combining K-means with LR, DNN and LSTM

The primary benefit of a combined model lies in its ability to leverage the strengths of its component techniques, resulting in a more robust learning pattern. By addressing the weaknesses of individual methods, the combined model can improve overall accuracy and reinforce the respective advantages [33] of the component methods. Our model utilizes deep learning and LSTM, a cutting-edge machine learning technique that utilizes artificial neural networks and has recently emerged as an approach for solar power forecasting [19]; this particular form of deep learning has been validated and showcased as a highly effective model for tackling numerous complex learning tasks, including load forecasting [22].

## 4. Model Results and Evaluations

### 4.1. Data Collection

After determining the optimum number of clusters using the elbow method, the meters were grouped into 4 clusters, based on their historical load generation and various weather factors. Figure 3 shows the clustered meters plot.



**Figure 3.** K-means clustering results.

### 4.2. Forecasting Methods Results

The performance evaluation of the three models utilized in this study was conducted by analyzing their respective metrics, including mean absolute error (MAE) and root mean squared error



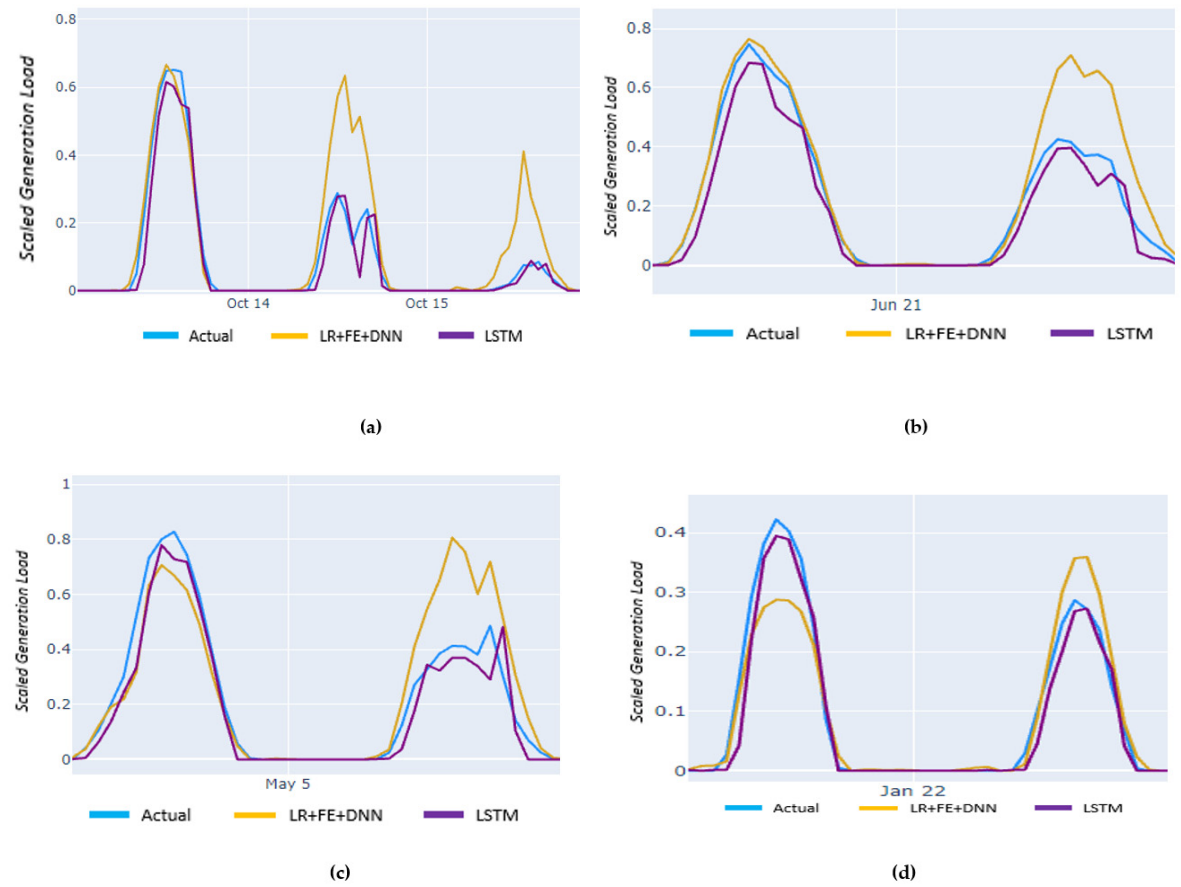
(RMSE), as presented in Table 2. The method results were compared with the baseline model provided by the utility, with MAE of 14.32%, which represented their existing model.

**Table 2.** Performance metrics evaluation.

Models.	MAE	RMSE
Utility Baseline	14.32%	23.13%
LR+DNN	4.81%	8.56%
LR+FE+DNN	2.49%	5.01%
LSTM	2.69%	4.68%

Among the models tested, the LR + FE + DNN model demonstrated the best MAE values, 2.49%, indicating strong average performance. However, it exhibited signs of overfitting, implying that it did not generalize well when presented with unseen data. On the other hand, the LSTM model showcased the lowest RMSE scores, 4.68%, suggesting superior overall performance when dealing with outliers or extreme values within the dataset. Moreover, the LSTM model displayed consistent performance across both the training and testing datasets, indicating its ability to effectively generalize and adapt to new data.

Notably, from Table 2 and Figure 4 (a) (b) (c) and (d), the LSTM model, which yielded the lowest RMSE, closely approximated the actual peak generation values compared with the alternative model. Also, the model has a better prediction when there is a change in the actual generation load from one day to another as seen in Figure 4 (a) (b) (c) and (d). Based on these observations, it can be concluded that the LSTM model outperformed the other two models in terms of predictive accuracy and reliability.



**Figure 4.** Forecasted outputs from models compared to actual generation. X axis is hourly/ Y axis is scaled generation in KW. (Generation Load Data is scaled between 0 and 1 due to customer privacy).

Overall, the findings highlight the favorable performance of the LSTM model and its potential for effectively forecasting generation load.

The evaluation of the LR + FE + DNN model unveiled a noteworthy concern regarding its tendency to overpredict generation load, particularly in scenarios involving daily fluctuations in the actual load generation. This persistent overprediction phenomenon indicates that the model consistently projected values that exceeded the true load values. Such a pattern can have adverse effects, including inefficient resource allocation and operational inefficiencies.

In contrast, the LSTM model demonstrated a superior capability to address this issue and offer more accurate load predictions. Leveraging its ability to capture long-term dependencies and effectively handle sequential data, the LSTM model proved adept at capturing and adapting to the daily changes in actual load generation.

The visual representations in Figure 4 (a) (b) (c) and (d), showcase the remarkable proximity of the LSTM model’s forecasted outputs to the actual peak generation levels, serving as evidence of its enhanced accuracy in predicting load variations. By successfully capturing the underlying patterns and dynamics of the data, the LSTM model effectively tackled the challenges posed by daily fluctuations, resulting in predictions that closely aligned with the true values.

The significance of avoiding overprediction of generation load lies in its practical implications. Overestimating load demand can lead to undesired outcomes, such as excessive resource allocation, unnecessary costs, and potential strain on the power grid. By effectively mitigating the overprediction issue, the LSTM model showcased its ability to deliver more reliable and precise load forecasts. This will enable utilities to optimize resource allocation, enhance operational efficiency, and maintain a stable and resilient power system.

To establish a benchmark, the results from other research from Table 1 were included in the comparison, as shown in Table 3 [23,25,26,29]–[31] exhibit a relatively high MAE and RMSE in comparison with other methods. These models may not accurately capture the variability and patterns in the data [24] and [28] achieve a modest MRE, indicating relatively accurate forecasts compared with others; these are combined methods showing higher accuracy than single methods. In comparison with the best models referenced in Table 3, our best models achieved better results, with MAE 2.49% and RMSE 4.68%. This is likely due to the fact that our method combines the advantages of K-means clustering and LSTM networks. K-means clustering can identify patterns in solar power generation data, while the LSTM model is able to determine the temporal relationships between the data [34].

**Table 3.** Comparison of performance metrics.

Ref #	Method Type	Error between Forecasted and Actual
[23]	Probabilistic	
[26]	Gradient Boosting trees	MAE > 5%
[30]	Gaussian Process Regression	RMSE > 6%
[31]	Autoregression	
[24]	K means Clustering with LSTM	MRE 2.52%
[25]	ARX	MAE 2%-6% RMSE >6%
[28]	SOM and Fuzzy	MRE 3.295%
[29]	SVR	RMSE 2.841%
[32]	NN Ensemble	MAPE >5%
This article best models	K-means Clustering with LR+FE+DNN	MAE 2.49% RMSE 5.01%
	K-means Clustering with LSTM	MAE 2.69% RMSE 4.68%

## 5. Conclusion and future work

The research conducted in this study is intended for electric utilities. It utilized its load generation, AMI, and weather data to forecast generation load using three different machine learning methods: linear regression, combined linear regression with deep neural networks, and long short-term memory (LSTM) forecasting. The results were compared using MAE and RMSE values in Tables 2 and 3.

The research revealed the indispensable role of weather features in achieving precise forecasts for generation load. Considering the substantial variations in weather patterns across different seasons, accounting for these features proved critical for accurate predictions. Among the models assessed, the combined approach of linear regression with DNN, incorporating weather features, emerged as the most effective in generating highly accurate predictions for generation load, exhibiting minimal error. However, this model tended to over-forecast when weather changes caused the generation to increase or decrease from one day to the next. On the other hand, the LSTM model was able to solve this problem with great accuracy and precision, resulting in the lowest RMSE scores and consistent performance on both the train and test datasets.

The success of these models opens possibilities for future work, such as monitoring solar locations and spotting unusual event behaviors. Additionally, the presented techniques can be applied to forecast the load of wind power generation. Overall, the findings suggest that incorporating weather data and using advanced machine learning techniques such as LSTM can significantly improve the accuracy and precision of generation load forecasting in the electric utility industry.

**Author Contributions:** Conceptualization, A.S.; methodology, A.S.; software, A.S.; validation, A.S.; formal analysis, A.S.; investigation, A.S.; resources, A.S.; data curation, A.S.; writing—original draft preparation, A.S.; writing—review and editing, W.S.; visualization, A.S.; supervision, W.S.; project administration, A.S.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability and Conflicts of Interest:** The data used in this research are confidential. The code and results are available. The authors declare no conflict of interest.

## References

1. M. J. E. Alam, K. M. Muttaqi, and D. Sutanto, "Effective utilization of available PEV battery capacity for mitigation of solar PV impact and grid support with integrated V2G functionality," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1562–1571, Oct. 2016, DOI: 10.1109/TSG.2015.2487514.
2. J. Shi, W. J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Trans. Ind. Appl.*, vol. 48, no. 3, pp. 1064–1069, Mar. 2012, DOI: 10.1109/TIA.2012.2190816.
3. U.S. Department of Energy, "Solar futures study fact sheet." Available: [https://www.energy.gov/sites/default/files/2021-09/Solar\\_Futures\\_Study\\_Fact\\_Sheet.pdf](https://www.energy.gov/sites/default/files/2021-09/Solar_Futures_Study_Fact_Sheet.pdf) Accessed: Apr. 15, 2023
4. S. Koochi-Kamali, N. A. Rahim, H. Mokhlis, and V. V. Tyagi, "Photovoltaic electricity generator dynamic modeling methods for smart grid applications: A review," *Renewable Sustain. Energy Rev.*, vol. 57, pp. 131–172, May 2016, DOI: 10.1016/j.rser.2015.12.137.
5. S. Sobri, S. Koochi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Convers. Manag.*, vol. 156, pp. 459–497, Jan. 2018, DOI: 10.1016/j.enconman.2017.11.019.
6. T. Hong *et al.*, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *Int. J. Forecasting*, vol. 32, no. 3, pp. 896–913, Jul. 2016, DOI: 10.1016/j.ijforecast.2016.02.001.
7. M. Sun, C. Feng, and J. Zhang, "Factoring behind-the-meter solar into load forecasting: Case studies under extreme weather," in *2020 IEEE Power Energy Soc. Innovative Smart Grid Technologies Conf. (ISGT)*, Washington, DC, 2020, pp. 1–5.
8. Z. Liu, Z. Zhu, J. Gao, and C. Xu, "Forecast methods for time series data: A survey," *IEEE Access*, vol. 9, pp. 91896–91912, Jun. 2021, DOI: 10.1109/ACCESS.2021.3091162.

9. K. G. Boroojeni *et al.*, "A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon," *Elect. Power Syst. Res.*, vol. 142, pp. 58–73, Jan. 2017, DOI: 10.1016/j.epsr.2016.08.031.
10. M. Madhukumar, A. Sebastian, X. Liang, M. Jamil, and M. N. S. K. Shabbir, "Regression model-based short-term load forecasting for university campus load," *IEEE Access*, vol. 10, pp. 8891–8905, Jan. 2022, DOI: 10.1109/ACCESS.2022.3144206.
11. S. Zhan, J. Wu, N. Han, J. Wen, and X. Fang, "Group low-rank representation-based discriminant linear regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 760–770, Feb. 2020, DOI: 10.1109/TCSVT.2019.2897072.
12. A. M. Kuriakose *et al.*, "Comparison of artificial neural network, linear regression and support vector machine for prediction of solar PV power," in *2020 IEEE Pune Sect. Int. Conf. (PuneCon)*, Pune, India, 2020, pp. 1–6.
13. M. Abuella and B. Chowdhury, "Solar power forecasting using support vector regression," *arXiv preprint arXiv:1703.09851*, 2016.
14. J. Kaur, A. Goyal, P. Handa, and N. Goel, "Solar power forecasting using ordinary least square based regression algorithms," in *2022 IEEE Delhi Sect. Conf. (DELCON)*, New Delhi, India, 2022, pp. 1–6.
15. A. M. Mikaeil, W. Hu, and S. B. Hussain, "A low-latency traffic estimation based TDM-PON mobile front-haul for small cell cloud-RAN employing feed-forward artificial neural network," in *2018 20th Int. Conf. Transparent Opt. Netw. (ICTON)*, Bucharest, Romania, 2018, pp. 1–4.
16. A. Alzahrani, P. Shamsi, C. Dagli, and M. Ferdowsi, "Solar irradiance forecasting using deep neural networks," *Procedia Comput. Sci.*, vol. 114, pp. 304–313, Jan. 2017, DOI: 10.1016/j.procs.2017.09.045.
17. Y. Yu, J. Cao, and J. Zhu, "An LSTM short-term solar irradiance forecasting under complicated weather conditions," *IEEE Access*, vol. 7, pp. 145651–145666, Oct. 2019, DOI: 10.1109/ACCESS.2019.2946057.
18. H. Zhou *et al.*, "Short-term photovoltaic power forecasting based on long short-term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78063–78074, Jun. 2019, DOI: 10.1109/ACCESS.2019.2923006.
19. G. Li *et al.*, "Photovoltaic power forecasting with a hybrid deep learning approach," *IEEE Access*, vol. 8, pp. 175871–175880, Sep. 2020, DOI: 10.1109/ACCESS.2020.3025860.
20. Z. Zhen *et al.*, "Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image," *IEEE Trans. Ind. Appl.*, vol. 56, no. 4, pp. 3385–3396, Apr. 2020, DOI: 10.1109/TIA.2020.2984617.
21. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, DOI: 10.1162/neco.1997.9.8.1735.
22. R. Zhang, M. Feng, W. Zhang, S. Lu, and F. Wang, "Forecast of solar energy production - A deep learning approach," in *2018 IEEE Int. Conf. Big Knowl. (ICBK)*, Singapore, 2018, pp. 73–82.
23. [27] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation — with application to solar energy," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3850–3863, Jan. 2016, DOI: 10.1109/TPWRS.2015.2502423.
24. M. S. Hossain and H. Mahmood, "Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast," *IEEE Access*, vol. 8, pp. 172524–172533, Sep. 2020, DOI: 10.1109/ACCESS.2020.3024901.
25. C. Yang, A. A. Thatte, and L. Xie, "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 104–112, Nov. 2015, DOI: 10.1109/TSTE.2014.2359974.
26. J. R. Andrade and R. J. Bessa, "Improving renewable energy forecasting with a grid of numerical weather predictions," *IEEE Trans. Sustain. Energy*, vol. 8, no. 4, pp. 1571–1580, Apr. 2017, DOI: 10.1109/TSTE.2017.2694340.
27. X. G. Agoua, R. Girard, and G. Kariniotakis, "Short-term spatio-temporal forecasting of photovoltaic power production," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 538–546, Apr. 2018, DOI: 10.1109/TSTE.2017.2747765.
28. H. T. Yang, C. M. Huang, Y. C. Huang, and Y. S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, Apr. 2014, DOI: 10.1109/TSTE.2014.2313600.

29. U. K. Das *et al.*, "Optimized support vector regression-based model for solar power generation forecasting on the basis of online weather reports," *IEEE Access*, vol. 10, pp. 15594–15604, Feb. 2022, DOI: 10.1109/ACCESS.2022.3148821.
30. H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, "Short-term solar power forecasting based on weighted Gaussian process regression," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 300–308, Jan. 2018, DOI: 10.1109/TIE.2017.2714127.
31. R. J. Bessa, A. Trindade, and V. Miranda, "Spatial-temporal solar power forecasting for smart grids," *IEEE Trans. Ind. Inform.*, vol. 11, no. 1, pp. 232–241, Feb. 2015, DOI: 10.1109/TII.2014.2365703.
32. M. Q. Raza, N. Mithulananthan, J. Li, K. Y. Lee, and H. B. Gooi, "An ensemble framework for day-ahead forecast of PV output power in smart grids," *IEEE Trans. Ind. Inform.*, vol. 15, no. 8, pp. 4624–4634, Aug. 2019, DOI: 10.1109/TII.2018.2882598.
33. Y. Ren, P. N. Suganthan, and N. Srikanth, "Ensemble methods for wind and solar power forecasting—A state-of-the-art review," *Renewable Sustain. Energy Rev.*, vol. 50, pp. 82–91, Oct. 2015, DOI: 10.1016/j.rser.2015.04.081.
34. Y. Wang, Y. Shen, S. Mao, X. Chen and H. Zou, "LASSO and LSTM Integrated Temporal Model for Short-Term Solar Intensity Forecasting," in *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2933–2944, April 2019, doi: 10.1109/JIOT.2018.2877510.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.