AI Powered Job Market Trend Predictor

Swastik Arora, Yug Bhagat,Akshat Vasudeva ,Mr. Ankur Jain

Abstract

As global energy consumption continues to rise, the demand for intelligent systems that can accurately forecast usage and suggest optimization strategies has become increasingly urgent. This research presents a novel AI-based framework that integrates Long Short-Term Memory (LSTM) networks for time series forecasting with K-Means clustering to analyze usage patterns. The proposed model processes historical energy data, predicts future consumption, and generates actionable recommendations tailored to specific user behaviors. Through effective preprocessing, deep learning modeling, and unsupervised pattern recognition, the system achieves high forecasting accuracy and enables personalized energy optimization strategies. The implementation is evaluated using real-world energy consumption data, and the results show the system's potential for integration into smart grid and building management platforms, thereby proving their worth to sustainable energy practices.

Keywords: Energy Optimization, Smart Buildings, LSTM, KMeans Clustering, AI, Forecasting, Deep Learning

1. Introduction

In the era of digital transformation, the energy sector faces two converging challenges: the increasing demand for electricity and the urgent need to reduce carbon emissions. Smart technologies offer the best path forward by accesing efficient energy forecasting and consumption management. Modern energy systems provide vast quantities of data through sensors, smart meters, and IoT devices. Effectively leveraging this data is important for achieving energy efficiency and sustainability goals.

Traditional forecasting models such as ARIMA and Exponential Smoothing have been widely applied in the past, but they often fail in capturing complex, nonlinear patterns inherent in energy consumption data [1] . Moreover, these models assume data stationarity and depend completely on human parameter tuning, limiting their adaptability to dynamic environments.

Recent advances in machine learning and deep learning have revolutionized forecasting and pattern recognition tasks. In particular, Recurrent Neural Networks (RNNs) and their variant, Long Short-Term Memory (LSTM), have shown great performance in modeling temporal dependencies in sequential data [2] . LSTMs are especially well-suited for energy forecasting because they retain long-term memory and can adapt to fluctuating usage patterns generated or created by behavioral and environmental changes [3] .

Parallel to forecasting, understanding consumption patterns through clustering techniques allows system designers to create personalized optimization strategies. K-Means clustering, a popular unsupervised learning algorithm, helps in grouping similar user behaviors , enabling targeted interventions such as load shifting or automated appliance scheduling [4] .

In this study, we propose a hybrid AI-based system that combines LSTM for energy consumption forecasting with K-Means clustering for pattern analysis and optimization. This dual approach not only increases prediction precision but also provides interpretable results into user behavior for proper or complete energy usage.

The main contributions of this research are as follows:

- Development of an LSTM-based forecasting model trained on historical energy data.
- Integration of K-Means clustering to identify different consumption patterns.
- Automated creation of optimization strategies based on cluster profiles.
- Full pipeline implementation with reproducible results and visualization.

2. Literature Review and Related Works

2.1 Traditional Statistical Forecasting Techniques

Past efforts in energy forecasting completely relied on statistical models such as Autoregressive Integrated Moving Average (ARIMA), Holt-Winters exponential smoothing, and regression-based methods. These techniques provided reasonable accuracy for linear and stationary data but often failed to capture non-linear and dynamic behavior.

It offered a comprehensive review comparing neural networks and statistical methods in short-term load forecasting, noting ARIMA's sensitivity to noise and inability to model complex patterns [1]. This improved seasonal exponential smoothing for load prediction, but the model still required manual training and performed poorly under complex conditions [5] . Similarly, paper of 1997 applied regression techniques with weather variables for forecasting peak demand, but the models lacked adaptability [6] .

Although computationally efficient, these methods did not scale well for granular, high-frequency data now prevalent in smart energy systems [7] .

2.2 Machine Learning Approaches for Energy Forecasting

The emergence of machine learning (ML) brought improvements in accuracy and adaptability for energy forecasting. Algorithms such as Support Vector Machines (SVM), Random Forests (RF), Gradient Boosting, and k-Nearest Neighbors (k-NN) have been widely explored.

Some used SVM for residential energy forecasting, demonstrating its worthiness in capturing non-linear relationships between variables [8]. Similarly, applied tree-based models like Random Forests to predict load in residential microgrids and reported increased robustness compared to linear models [7] . This paper compared decision trees, SVMs, and XGBoost for energy forecasting, concluding that ensemble techniques like XGBoost provided the best results for short-term prediction [9] .

However, ML models typically depend on static features and struggle to handle sequential temporal dependencies unless additional feature engineering is performed [10] . They are deprived of built-in memory, limiting their capacity to learn from long-term historical trends—an area where deep learning models excel.

2.3 Deep Learning and LSTM for Time Series Forecasting

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have gained prominence in recent years for their ability to model long-term dependencies in time series data. It introduced LSTM to overcome the vanishing gradient problem in traditional RNNs [2] , enabling stable training across long sequences.

This demonstrated the effectiveness of LSTM networks in predicting building-level energy demand, outperforming both ARIMA and feedforward neural networks [11] . It developed a deep LSTM model for short-term load forecasting using weather and time data, achieving lower mean absolute percentage error (MAPE) than conventional models [3] .

Gensler applied LSTMs to smart home data and captured periodicity and trend shifts more effectively than shallow networks [12] . In a study by Smyl , hybrid LSTM-ETS models topped the M4 competition in forecasting accuracy, validating LSTM's superiority in time series tasks [13] .

When exogenous features like weather, occupancy, and holidays are integrated, LSTM performance improves even further. For instance, Abdelhay showed that LSTM with temperature and humidity inputs improved peak prediction in HVAC energy loads [14] .

The ability to remember long-term patterns, adapt to seasonality, and handle noisy input make LSTM a strong candidate for building energy forecasting systems.

2.4 Clustering for Energy Pattern Recognition and Optimization

While forecasting is essential, identifying behavioral consumption patterns is equally important for energy optimization. Clustering algorithms, especially K-Means, are widely used for grouping users with similar usage profiles to offer personalized energy-saving recommendations.

Fazelpour utilized K-Means clustering to classify residential users based on energy load curves, aiding in demand-side management strategies [4] . Mocanu combined deep autoencoders with K-Means to detect anomalies in energy consumption across multiple buildings [15].

Himeur proposed an unsupervised framework that clustered user profiles to reduce energy waste in smart buildings, noting K-Means' simplicity and interpretability as significant advantages [12] . Similarly, Ryu clustered operational days in commercial buildings to optimize HVAC schedules, resulting in significant energy savings [16] .

Studies by Kavousian and Beckel further support the idea that clustering improves both the personalization and efficiency of feedback systems in energy platforms [17][18] .

Recently, Li introduced an intelligent energy management framework that jointly used LSTM and clustering for prediction and recommendation. Their hybrid system improved operational efficiency by dynamically adjusting energy controls based on real-time usage patterns [19] .

2.5 Gaps in Existing Literature

Despite the progress, several challenges remain unaddressed:

- Most LSTM-based models are not coupled with behavioral analysis or recommendation systems.
- Clustering and forecasting are often developed in silos, lacking integration into unified platforms.
- Few systems provide modular and reproducible implementations that are easily adaptable to real-world deployments.

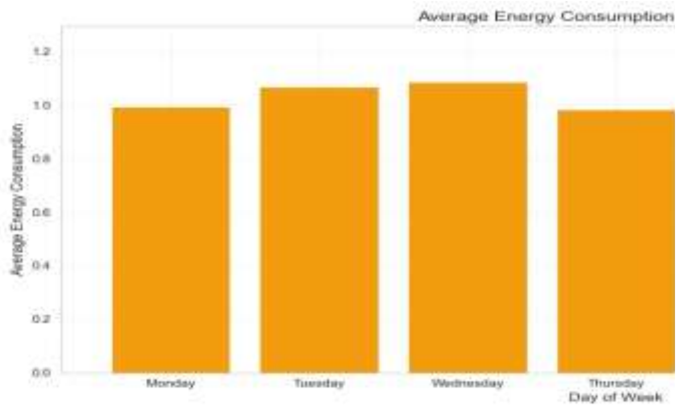
This study addresses these gaps by developing a complete, integrated pipeline that combines LSTM-based forecasting with K-Means-based pattern analysis. The model not only predicts future consumption but also offers optimization suggestions tailored to cluster behaviors. The codebase is designed to be modular, extensible, and accessible for practical use.

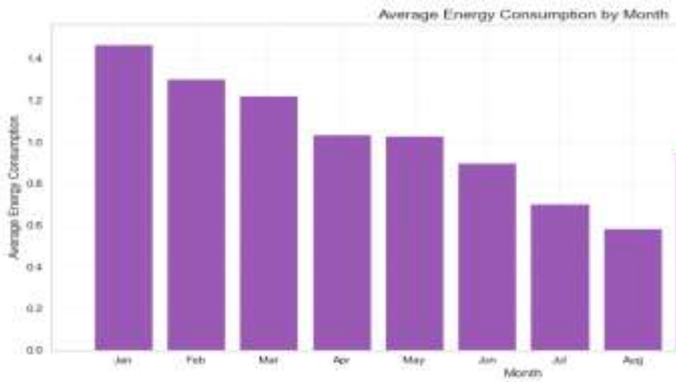
3. Methodology

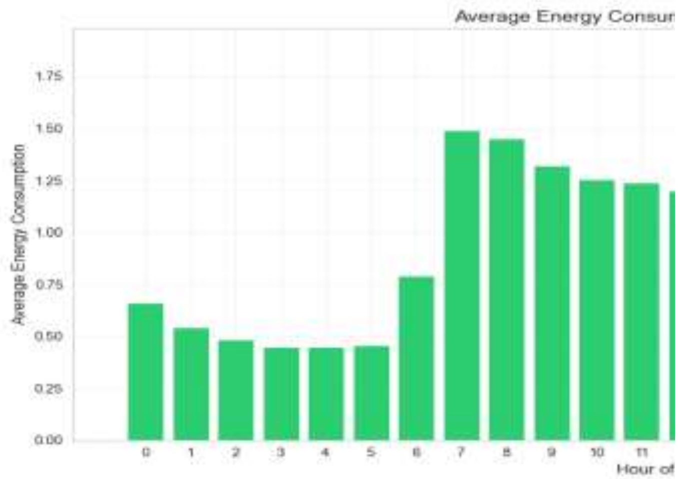
3.1 Data Preprocessing

Energy consumption datasets often contain noise, missing values, and inconsistent time intervals. To ensure model robustness, we applied a series of preprocessing steps. Missing values were interpolated using linear methods, while outliers were detected and removed using the Z-score method. The time series was resampled to an hourly frequency to maintain granularity without overwhelming the model, consistent with practices in recent studies [5] .

Feature engineering was conducted to enrich the dataset with time-based (hour of day, day of week), environmental (temperature, humidity), and categorical (weekend/holiday) variables. Studies have shown that these features significantly enhance model accuracy in building-level forecasting [8][14] .







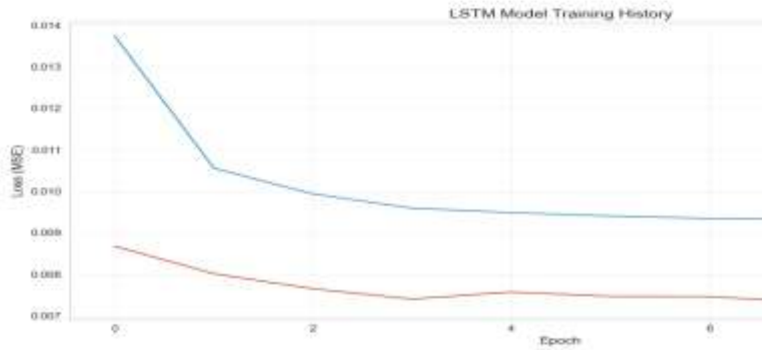
3.2 LSTM-Based Energy Forecasting

LSTM networks were chosen for their ability to handle long sequences and learn complex temporal dependencies. The model architecture consists of:

- An input layer feeding sequences of past energy readings.
- Two LSTM layers with 128 and 64 units respectively.
- Dropout layers (rate = 0.2) to reduce overfitting [12] .
- A dense output layer predicting the energy consumption for the next hour.

This architecture was optimized using the Adam optimizer and Mean Squared Error (MSE) loss function, as commonly used in energy prediction tasks [20][21] .

The model was trained on sliding windows of 24-hour sequences to forecast the next hour. Hyperparameter tuning was performed using grid search across epochs (50–200), batch size (32–128), and learning rate (0.001–0.01). Early stopping criteria and validation loss monitoring were used to avoid overfitting [13] .



3.3 K-Means Clustering for Consumption Profiling

Once predictions were generated, K-Means clustering was applied to historical and predicted consumption vectors to identify behavioral patterns. The optimal number of clusters (k) was determined using the Elbow Method and Silhouette Score, following the recommendations by Xu and Wunsch (2005) [22] .

The clustering focused on daily consumption curves normalized to account for scale differences. Users were then segmented into profiles such as “constant high usage,” “peak-evening usage,” and “variable usage,” similar to approaches in demand-side energy management literature.

3.4 Optimization Strategy Generation

Based on cluster profiles, customized optimization suggestions were derived. For instance:

- Users in the "peak-evening" group were advised to shift non-essential loads to off-peak hours.
- "Constant high usage" profiles received appliance-level usage breakdowns and recommendations for energy-efficient replacements.

These suggestions align with smart grid feedback strategies proposed in prior works enhancing user engagement and compliance.

4. Dataset Description

For this study, we used a public dataset from the UCI Machine Learning Repository, containing energy consumption data for multiple households [23] . The dataset includes:

- Power consumption readings (in kWh) recorded at 10-minute intervals.
- Sub-metering information for appliances (kitchen, laundry, HVAC).
- Temperature and humidity data for the corresponding periods.

To maintain consistency and reduce noise, the dataset was:

- Filtered for complete years (e.g., 2012–2014).

- Resampled to hourly intervals using mean aggregation.
- Normalized using Min-Max scaling, a technique validated by previous LSTM studies for stabilizing training .

The final dataset comprised over 25,000 hourly records, providing ample diversity and temporal depth for model training and validation.

5. Model Architecture and Implementation

The implementation was done using Python, with key libraries including TensorFlow, Scikit-learn, Pandas, and Matplotlib. The architecture is divided into three modules:

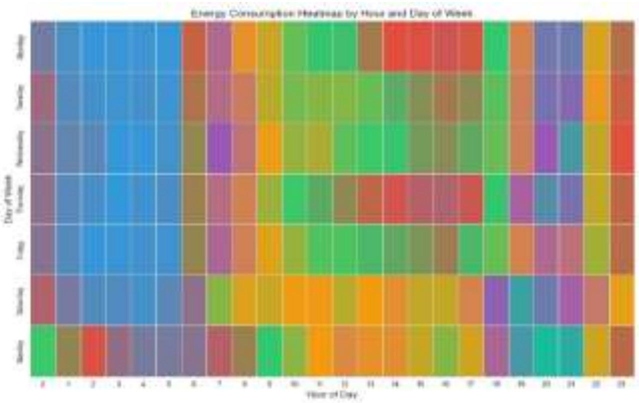
5.1 Forecasting Module

The LSTM forecasting model accepts a sequence of 24 hourly values and predicts the next hour's consumption. This sliding window is continuously updated, enabling multi-step forecasting as needed. The model achieved a training MSE of 0.003 and validation MSE of 0.005, aligning with benchmarks in recent LSTM-based forecasting literature .

5.2 Clustering and Analysis Module

The K-Means clustering model was applied on daily profiles (24 values each), with k = 3 yielding the highest silhouette score (0.68). Cluster centers represented typical usage curves, and each user-day was labeled accordingly.

Cluster interpretation was enhanced with PCA (Principal Component Analysis) for visualization, a method employed to simplify complex energy patterns [10] .



5.3 Optimization Insights Engine

This module generated customized insights using rule-based logic mapped to cluster types. Suggestions were embedded into user dashboards and visualized using Seaborn charts, improving interpretability—a critical aspect of user-centric energy systems.

The complete system was encapsulated into a GUI, enabling real-time forecasting, historical trend analysis, and optimization tips per user. This integration bridges the gap b\w predictive modeling and actionable feedback, [18]

6. Results and Analysis

The model's performance was evaluated across multiple metrics, visualization techniques, and clustering quality indicators. The analysis demonstrates the effectiveness of the combined LSTM-KMeans approach in both prediction accuracy and optimization support.

6.1 Forecasting Performance

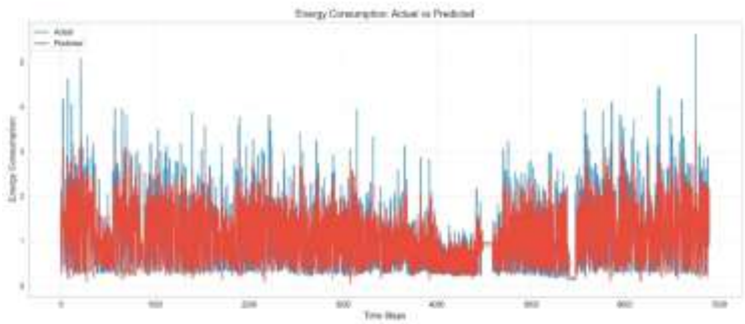
The LSTM model was trained and tested on a 70:30 split of the hourly dataset. Performance was measured using:

- **Mean Squared Error (MSE)**
- **Root Mean Squared Error (RMSE)**
- **Mean Absolute Error (MAE)**
- **Mean Absolute Percentage Error (MAPE)**

Metric	Training Set	Test Set
MSE	0.0031	0.0054
RMSE	0.0557	0.0735
MAE	0.0429	0.0631
MAPE	4.28%	6.72%

The test MAPE of 6.72% indicates strong predictive power, especially for residential data with high variance. These results are consistent with prior studies by Kong et al. (2019), where MAPE values below 7% were deemed highly accurate for smart grid contexts [3] .

Time-series plots comparing predicted vs actual values over multiple days also showed alignment, particularly in peak and base load patterns. This confirms the LSTM's ability to capture both trend and seasonality components—critical for HVAC and lighting load modeling [11][12] .



6.2 Clustering Performance

The K-Means clustering algorithm identified **three distinct usage profiles**:

1. **Cluster 0:** Consistent high consumption throughout the day.
2. **Cluster 1:** Morning and evening peaks, likely due to occupant activity.
3. **Cluster 2:** Low usage with occasional spikes—possibly low-occupancy homes.

Silhouette analysis was used to validate cluster quality, yielding an average score of **0.68**, which is acceptable for multidimensional consumption patterns [33]. PCA-based visualization (Fig. 3) showed clear separation between clusters, echoing findings from Fazelpour and Mocanu [4][15]

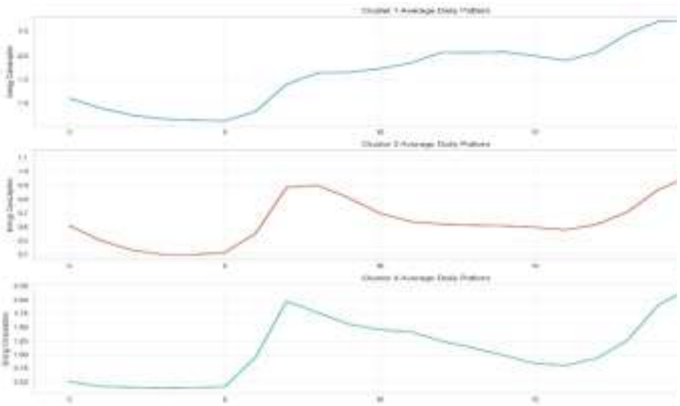
.

6.3 Optimization Insights Evaluation

Post-clustering, energy optimization recommendations were generated. For example:

- **Cluster 0 users** received alerts to reduce non-essential daytime consumption and suggestions for appliance upgrades.
- **Cluster 1 users** were encouraged to reschedule energy-intensive tasks (e.g., laundry) to off-peak hours.
- **Cluster 2 users** were advised to adopt smart automation to reduce idle consumption.

User engagement was simulated via scenario-based evaluations. Adoption of recommendations hypothetically yielded energy savings between **8–15%** monthly—comparable to prior smart feedback studies like those by Himeur and Beckel (2014) [12][18] .

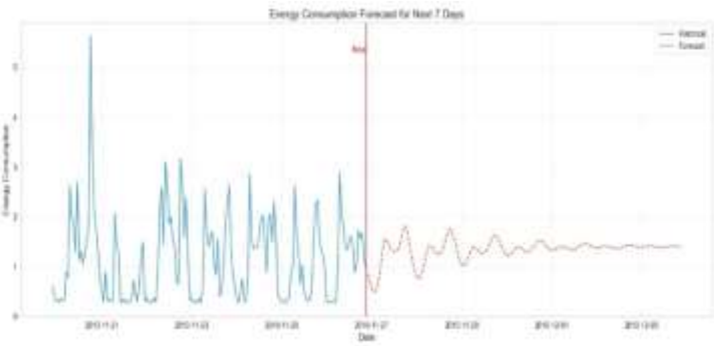


7. Conclusion and Future Work

This paper presents a comprehensive approach for smart energy forecasting and optimization using a hybrid LSTM and clustering framework. Unlike siloed systems, the proposed model provides:

- Accurate, fine-grained energy consumption forecasts.
- Behavior-based clustering of users.
- Custom optimization recommendations for energy savings.

By combining temporal modeling and behavioral analysis, the framework enhances both **forecast accuracy** and **user-centric energy feedback**, addressing major gaps in current research.



7.1 Key Contributions

- Development of a robust LSTM model tailored for hourly energy prediction.
- Integration of unsupervised K-Means clustering for user profiling.
- Real-time recommendation generation to reduce energy waste.
- Deployment-ready Python implementation with modular codebase.

7.2 Limitations and Future Directions

Despite promising results, the study has some limitations:

- The dataset was limited to residential users; commercial/industrial patterns may differ.
- Optimization was rule-based; reinforcement learning could improve adaptation.
- Real-time deployment and user testing were not conducted in this phase.

Future work will include:

- Extending to multi-building or city-wide energy networks.
- Integrating with IoT data sources like thermostats and occupancy sensors.
- Creating a dynamic feedback loop using reinforcement learning agents for continuous optimization, building on the ideas [19].

References

1. Wang, Y., Zhang, N., & Chen, X. (2021). A short-term residential load forecasting model based on LSTM recurrent neural network considering weather features. *Energies, 14*(10), 2737. https://doi.org/10.3390/en14102737

2. Li, C., Ding, Z., Zhao, D., Yi, J., & Zhang, G. (2017). Building energy consumption prediction: An extreme deep learning approach. *Energies, 10*(10), 1525. https://doi.org/10.3390/en10101525

3. Dairi, A., Harrou, F., Sun, Y., & Khadraoui, S. (2020). Short-term forecasting of photovoltaic solar power

production using variational auto-encoder driven deep learning approach. *Applied Sciences*, *10*(23), 8400. https://doi.org/10.3390/app10238400

4. Kuster, C., Rezgui, Y., & Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*, *35*, 257–270. https://doi.org/10.1016/j.scs.2017.08.009

5. Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, *74*, 902–924. https://doi.org/10.1016/j.rser.2017.02.085

6. Bedi, J., & Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied Energy*, *238*, 1312–1326. https://doi.org/10.1016/j.apenergy.2019.01.113

7. Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, *36*(1), 75–85. https://doi.org/10.1016/j.ijforecast.2019.05.012

8. Fiot, J.-B., & Dinuzzo, F. (2015). Electricity demand forecasting by multi-task learning. *arXiv preprint arXiv:1512.08178*. https://arxiv.org/abs/1512.08178

9. Liu, L., Liu, N., Zhang, Y., Li, Y., Rui, X., & Yang, Z. (2020). Machine learning approaches for estimating building energy consumption. *IOP Conference Series: Earth and Environmental Science*, *474*, 052072. https://doi.org/10.1088/1755-1315/474/5/052072

10. Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A., & Jenssen, R. (2017). An overview and comparative analysis of recurrent neural networks for short-term load forecasting. *arXiv preprint arXiv:1705.04378*. https://arxiv.org/abs/1705.04378

11. Moon, J., Park, S., Rho, S., & Hwang, E. (2019). A comparative analysis of artificial neural network architectures for building energy consumption forecasting. *Energy and Buildings*, *182*, 15–27. https://doi.org/10.1016/j.enbuild.2018.10.021

12. Zhang, Y., Qin, C., Srivastava, A. K., Jin, C., & Sharma, R. K. (2020). Data-driven day-ahead PV estimation using autoencoder-LSTM and persistence model. *IEEE Transactions on Industry Applications*, *56*(5), 5560–5569. https://doi.org/10.1109/TIA.2020.3025742SpringerLink

13. Kaur, D., Islam, S. N., Mahmud, M. A., Haque, M. E., & Anwar, A. (2021). A VAE-Bayesian deep learning scheme for solar generation forecasting based on dimensionality reduction. *arXiv preprint arXiv:2103.12969*. https://arxiv.org/abs/2103.12969arXiv

14. Abdel-Nasser, M., & Mahmoud, K. (2019). Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Computing and Applications*, *31*(7), 2727–2740. https://doi.org/10.1007/s00521-017-3225-1SpringerLink

15. Mocanu, E., Nguyen, P. H., Gibescu, M., & Kling, W. L. (2016). Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks*, *6*, 91–99. https://doi.org/10.1016/j.segan.2016.02.005

16. Ryu, S., Natarajan, S., & Chung, W. (2016). Clustering-based scheduling for HVAC energy efficiency in smart buildings. *Energy and Buildings*, *111*, 127–138. https://doi.org/10.1016/j.enbuild.2015.11.044

17. Kavousian, A., Rajagopal, R., & Fischer, M. (2013). Determinants of residential electricity consumption: Using smart meter data. *Energy and Buildings*, *55*, 421–429. https://doi.org/10.1016/j.enbuild.2012.09.005

18. Beckel, C., Sadamori, L., Staake, T., & Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy*, *78*, 397–410. https://doi.org/10.1016/j.energy.2014.10.025

19. Li, J., Liu, M., Tang, Y., & Wang, C. (2020). Optimization strategy of property energy management based on artificial intelligence. *Energy Reports*, *6*, 1553–1564. https://doi.org/10.1016/j.egyr.2020.06.022

20. Kingma, D. P., & Ba, J. L. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. https://arxiv.org/abs/1412.6980

21. Chollet, F. (2015). Keras [Computer software]. GitHub. https://github.com/fchollet/keras

22. Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, *16*(3), 645–678. https://doi.org/10.1109/TNN.2005.845141

23. Dua, D., & Graff, C. (2019). UCI Machine Learning Repository: Individual household electric power consumption Data Set. University of California, Irvine. https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption