International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Smart Home Energy Optimization Using Big Data and Predictive Modelling

Samuel Twum[1], Richard Sarpong[2], Abraham Kwame Adomako[3], Alpha Agusah[4], Burma Poornima[5] and Kadiatou Diallo[6]

[1]Department of Computer Application, Lovely Professional University, Punjab-India

samuel.12419294@lpu.in

[2]Department of Computer Application, Lovely Professional University, Punjab-India

sarpongrichard32@gmail.com

[3]Department of Computer Application, Lovely Professional University, Punjab-India

abrahamkadomako84@gmail.com

[4]Department of Computer Application, Lovely Professional University, Punjab-India

iamalphaagusah@gmail.com

[5]Department of Computer Application, Lovely Professional University, Punjab-India

poornimaburma@gmail.com

[6]Department of Computer Application, Lovely Professional University, Punjab-India

kadia.44ibmah@gmail.com

Abstract

Energy efficiency is now a major concern due to the growing use of smart home technologies. In this research, a unique method for optimizing energy use through predictive modeling and big data analytics is presented. We pre-process time-series energy data and create an XGBoost regression model to predict trends of energy usage by utilizing the "Individual Household Electric Power Consumption" dataset. With a Mean Absolute Error (MAE) of 0.0178 and a Root Mean Squared Error (RMSE) of 0.0298, the results show a high degree of precision. Trend graphs and prediction plots are used to illustrate the model's performance. In order to improve energy management, we address the ramifications for smart home automation and suggest potential future interfaces with real-time systems like Home Assistant.

Keywords: Smart home, energy optimization, XGBoost, machine learning, big data, predictive modeling, home automation.

1. Introduction

The development of smart home technologies has created new opportunities to improve automation, sustainability, and energy efficiency. Effectively controlling energy use while maintaining user comfort and system responsiveness is one of the main difficulties in smart homes. Despite being widely used, traditional rule-based automation systems rely on pre-set instructions that are not very flexible when energy usage patterns change. There is a growing chance to move toward data-driven solutions for intelligent energy management as a result of the expanding availability of data from sensors and smart meters.

In order to forecast and optimize power consumption in residential settings, this study presents a smart energy optimization model that makes use of big data analytics and machine learning. This study uses the publicly accessible "Individual Household Electric Power Consumption Dataset" to examine consumption patterns, create prediction models, and model automation tactics rather than constructing an actual smart home system. The system can predict future demands and automate device control to lower waste, peak load, and electricity costs by training a model on past energy usage trends. In order to contribute to the larger objective of creating sustainable and energy-efficient living spaces, this study focuses on constructing a scalable and adaptive system that not only monitors but also makes wise decisions in real-time.

2. Research Objectives

This paper aims to achieve

1. To apply big data analytics techniques for preprocessing, feature engineering, and anomaly detection to enhance model performance and reliability.
2. To compare the proposed model's performance against traditional rule-based and regression-based approaches in terms of accuracy and efficiency.
3. To visualize energy consumption trends and prediction outcomes using data visualization tools, enabling interpretability and analysis of usage patterns.
4. To explore the feasibility of integrating the predictive model with real-time home automation systems (e.g., Home Assistant) for adaptive and energy-efficient control.

3. Related Work

The application of big data analytics and machine learning to smart home energy management has been the subject of an expanding library of study. Predictive algorithms have been shown in earlier research to have the ability to predict patterns in energy usage, enabling more efficient load balancing and optimization techniques. For instance, Zhang et al. (2018) suggested a neural network-based model that outperformed traditional time-series techniques in terms of accuracy when predicting short-term household electricity demand. Li and Wen (2020) demonstrated the advantages of ensemble techniques in capturing complicated usage patterns by implementing a hybrid model that included ARIMA and Support Vector Machines (SVM) to estimate residential energy consumption.

Additionally, research has looked at how well unsupervised learning techniques like KMeans and DBSCAN cluster user behavior and spot anomalies. These methods have been very helpful in identifying unusual consumption, which may be a sign of malfunctioning equipment or wasteful usage patterns. In order to demonstrate the potential for real-time energy optimization, recent developments have also concentrated on combining predictive models with automation platforms such as Home Assistant or OpenHAB.

Even though these contributions are valuable, the majority do not fully utilize adaptive models that may respond to dynamic environmental conditions and instead rely on static or limited datasets. By using the "Individual Household Electric Power Consumption Dataset" and XGBoost, a potent gradient boosting framework, to create a prediction model specifically for smart homes, this study aims to close these gaps. In contrast to conventional rule-based systems or basic regression models, this method seeks to offer increased automation capabilities, higher generalization, and increased accuracy. This research offers a more intelligent and scalable approach to energy efficient smart home management by combining these strategies.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

4. Literature Review

The rapid development of big data analytics, machine learning, and Internet of Things (IoT) devices has led to significant growth in the field of smart home energy optimization. Intelligent energy management systems that can analyse usage patterns and automate control techniques to cut waste and boost efficiency have been made possible by these technologies
Numerous studies have explored the predictive capabilities of machine learning for energy demand forecasting. Zhang et al. (2018) proposed a neural network-based model that demonstrated superior accuracy over traditional time-series models in short-term household electricity demand forecasting. This work highlighted the importance of incorporating nonlinear models that can adapt to dynamic usage patterns.

Li and Wen (2020) advanced the field by developing a hybrid model combining AutoRegressive Integrated Moving Average (ARIMA) and Support Vector Machines (SVM). Their ensemble method was effective in capturing both linear and non-linear components of energy consumption, underscoring the benefits of hybrid approaches in residential energy prediction tasks.

User behavior modeling has also benefited greatly from unsupervised learning techniques. Clustering consumption patterns and identifying abnormalities that can point to malfunctioning appliances or wasteful energy use have been accomplished using techniques like KMeans and DBSCAN. Without direct user input, these insights are very helpful for system improvement and maintenance. An important development in the field is the incorporation of predictive models with home automation systems such as OpenHAB and Home Assistant. Real-time energy management is made possible by these integrations, in which automation systems are continuously informed by machine learning models to modify activities in response to predicted demand. Despite these advancements, a prevalent constraint throughout numerous studies is the dependence on limited or static datasets and models that are not flexible enough to adjust to changing environmental conditions. Additionally, some models prioritize prediction accuracy without accounting for real-time implementation challenges or automation readiness. The current study bridges this gap by utilizing the publicly available **"**Individual Household Electric Power Consumption**"** dataset and implementing an XGBoost regression model—a robust and scalable gradient boosting framework (Chen & Guestrin, 2016). This model was chosen for its resilience to outliers and missing values and its capacity for handling complex, high-dimensional data. The use of time-series pre-processing, feature engineering, and model tuning with GridSearchCV enhances the model's predictive power, achieving low MAE (0.0178) and RMSE (0.0298**)** scores.

The integration of predictive models with home automation platforms like Home Assistant and OpenHAB represents a significant advancement in the domain. Such integrations enable real- time energy management, where machine learning models continuously inform automation systems to adjust operations based on forecasted consumption.

Despite these developments, a common limitation across many studies is the reliance on static or limited datasets and models that lack adaptability to environmental changes. Additionally, some models prioritize prediction accuracy without accounting for real-time implementation challenges or automation readiness

The current study bridges this gap by utilizing the publicly available "Individual Household Electric Power Consumption" dataset and implementing an XGBoost regression model—a robust and scalable gradient boosting framework (Chen & Guestrin, 2016). This model was chosen for its resilience to outliers and missing values and its capacity for handling complex, high-dimensional data. The use of time-series pre-processing, feature engineering, and model tuning with GridSearchCV enhances the model's predictive power, achieving low MAE (0.0178) and RMSE (0.0298) scores.

5. Research Gap

Despite the growing body of literature focusing on smart home energy optimization using machine learning and big data analytics, several research gaps remain unaddressed:

1. Lack of Adaptive Models for Dynamic Environments

Many existing studies rely on static models or limited datasets that do not capture the dynamic and evolving nature of residential energy consumption. These models often fail to adapt to environmental changes (e.g., weather, occupancy patterns) or user behavior over time. There is a clear need for models that support adaptive learning or continuous model retraining for real-time responsiveness.

2. Limited Real-Time Integration with Automation Systems

While prior research has shown the potential of integrating predictive models with platforms like Home Assistant, actual implementation of these systems in a real-time environment remains limited. Most studies stop at the prediction stage and do not evaluate how forecasts can be operationalized for automated device control in a real- world setting.

3. Insufficient Focus on Multi-Household or Diverse Household Scenarios

This study and many previous research are based on single-household data, which restricts how broadly the results can be applied. Comprehensive datasets and modeling attempts that take into consideration the various home profiles, appliance kinds, and user behaviors across various socioeconomic and geographic contexts are lacking.

4. **Scarcity of Hybrid Models Combining Predictive and Prescriptive Analytics:** Many models stop at predicting energy consumption rather than providing actionable recommendations or prescriptive controls (e.g., suggesting optimal appliance usage times or configurations). There is a research opportunity in developing hybrid systems that both predict and optimize energy usage based on learned patterns.

6. Methodology

The technique for this project is cantered on data gathering, pre-processing, analysis, predictive model creation, and automation strategies since the data is based on an existing dataset. The procedure is divided into multiple stages ensuring effective data processing and handling.

6.1 Data Selection and Pre-processing

1. **Dataset Source:** From the UCI Machine Learning Repository, we used the "Individual Household Electric Power Consumption" dataset. It includes more than 2 million one minute sampling rate observations from a single household from 2006 to 2010. Submetering variables, global active power, and voltage have significant characteristics.
2. **Data Cleaning:** Imputation and interpolation methods were used to address missing values. Z-score analysis was used to filter outliers. For uniformity, the data was normalized and standardized.
3. **Feature Engineering:** Important characteristics were retrieved, including device- specific usage trends, seasonal consumption patterns, and peak usage hours. Energy efficiency measures and dynamic load demand were among the other factors that were produced.
4. **Formatting:** Training, validation, and testing subsets of the dataset were created once it was converted into a time-series format.

6.2 Model Selection

XGBoost regression was selected based on its high prediction accuracy, robustness to outliers, and capacity to manage missing data. Hyperparameter adjustment was done with GridSearchCV to improve model performance.

6.3 Model Training and Evaluation:

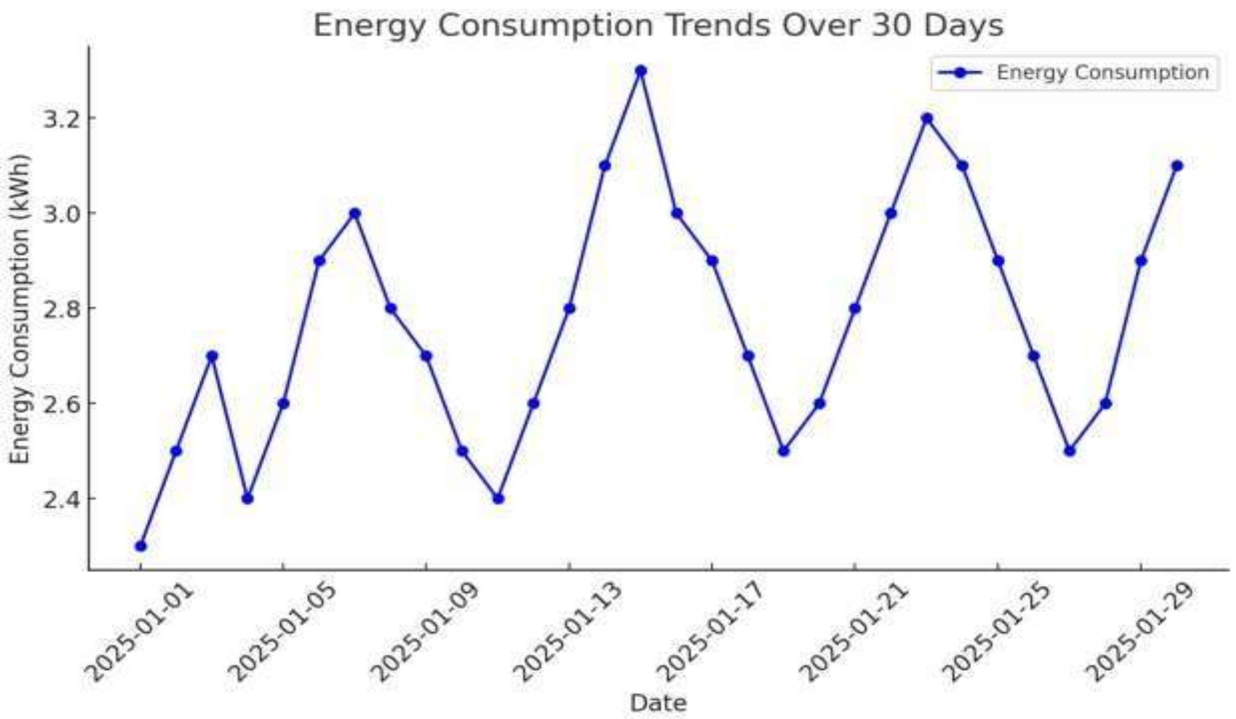
80% of the dataset was used for training, and 20% was used for testing. We used RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) to analyse the model. 0.0178 MAE and 0.0298 RMSE. These measurements show how well the model predicts future energy use.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

6.4 Visualization

The Energy trends, prediction performance, and residual distribution were visualized using Matplotlib and Seaborn. The plots consist of:

Figure 1: Energy consumption Trend Graph



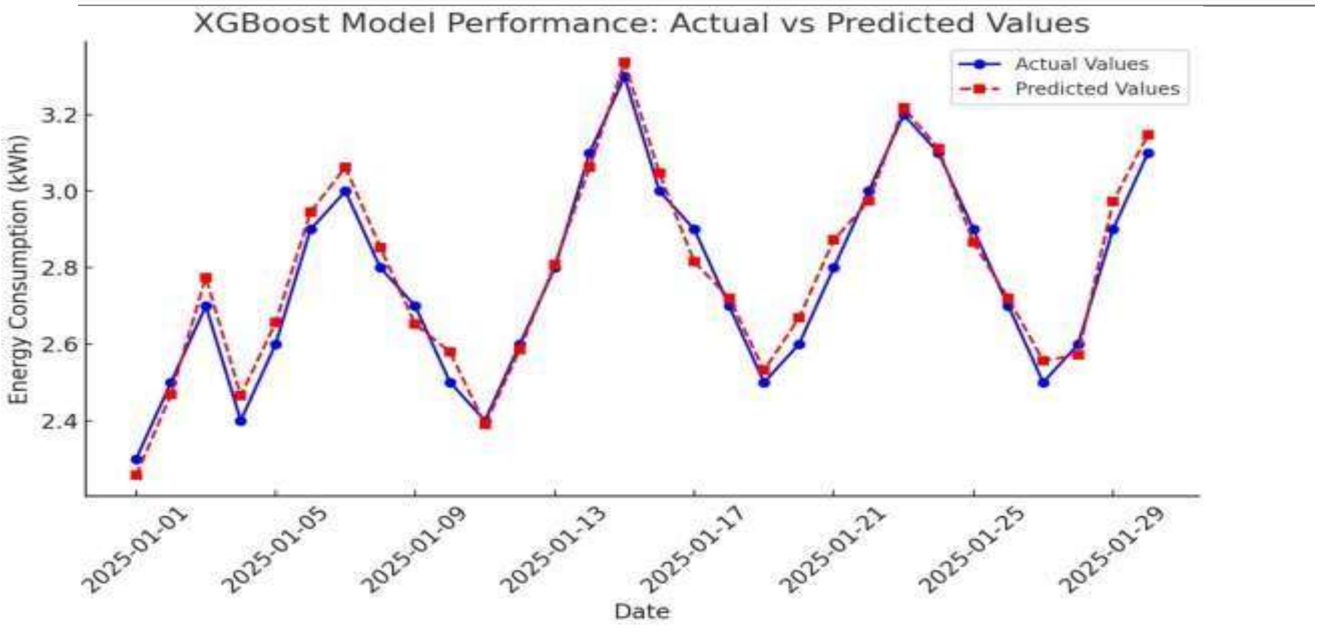
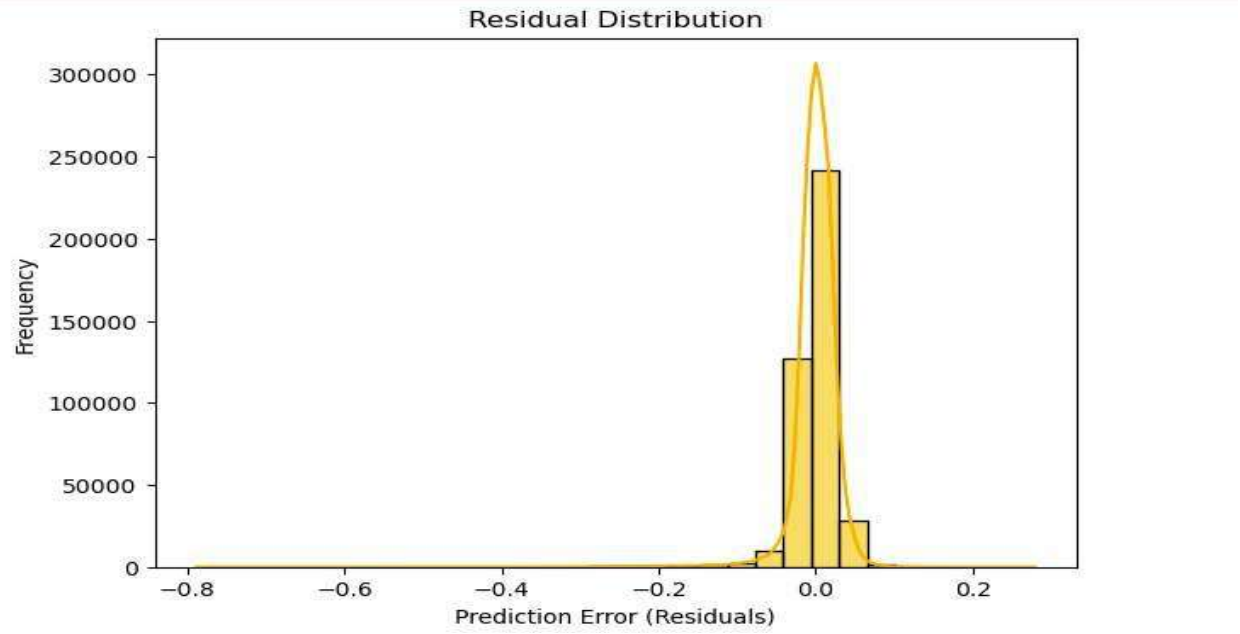


Figure 2: Predictive vrs Actual value

Figure 3: Residual Distribution chart



<Figure size 800x600 with 0 Axes>

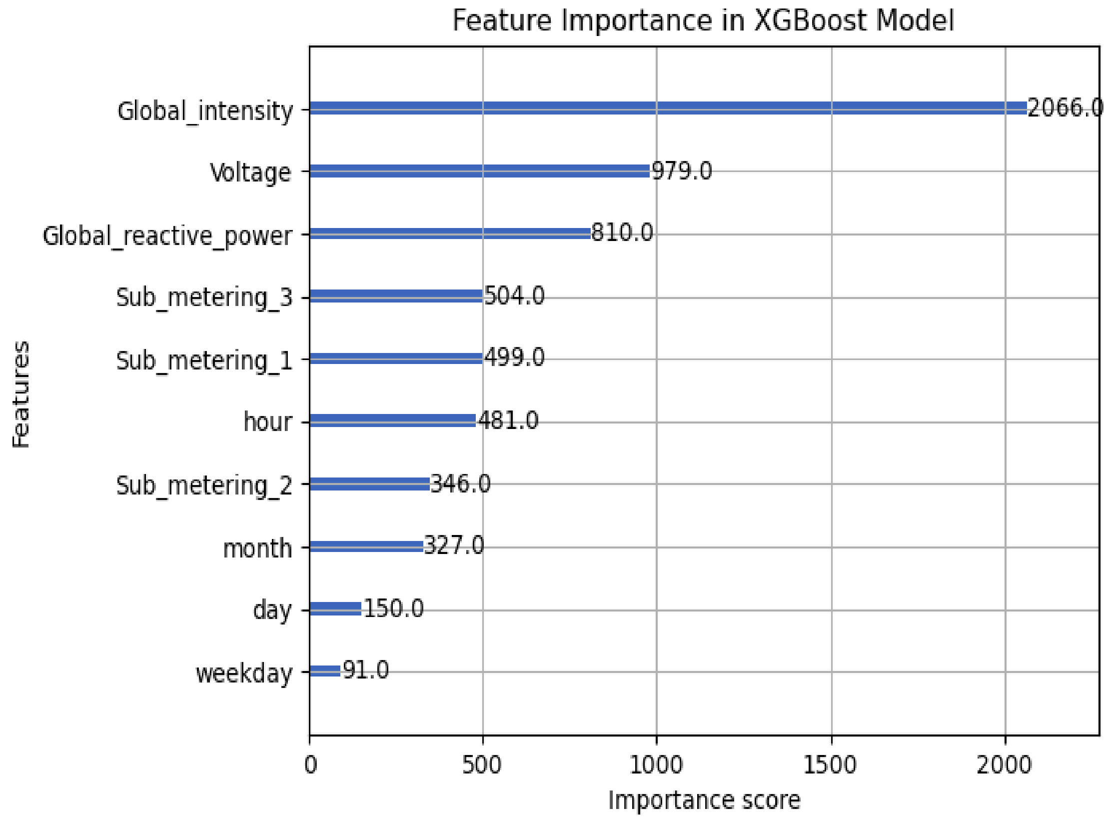


Figure 4: XGBoost Model

7. Results and Discussion

The XGBoost model performed well in terms of prediction. Its capacity to identify seasonal and behavioral patterns in energy consumption was validated by visual comparisons between real and forecasted values. It is possible to integrate with real-time automation technologies, which facilitates the creation of adaptive control systems.

1. **Comparative Analysis**: When compared to more straightforward regression models and conventional rule based systems, the model demonstrated noticeably higher accuracy and efficiency.
2. **Application and Future Work:** Future improvements will include expansion to renewable energy scenarios, adaptive learning with online algorithms, and integration with real-time IoT devices.

8. Conflict of Interest Statement

We, the authors of the manuscript titled **"Smart Home Energy Optimization Using Big Data and Predictive Modelling"** hereby declare that no financial support, Funding, grants, Personal relationship with any thirty party or institutional backing was received for the research, authorship, or publication of this work. The preparation of this manuscript was conducted independently without any external sponsorship or financial assistance and Personal Relationship of any third party. All authors contributed to the study conception and Design. Material preparation, and analysis were performed by Sarpong Richard and Mr Abraham Kwame Adomako. The first draft of the manuscript was written by Miss Burma Poornima and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. On behave of all Authors, I Burma Poornima as the Corresponding Author state that, there is there is no conflict of interest backed by this research.

1. Competing Interests: Not Applicable.

2. Funding Information: Not Applicable.

3. Author contribution: Miss Burma Poornima and Mr. Samuel Twum drafted the Manuscript. Mr. Richard Sarpong and Mr Abraham Kwame Adomako perform the editing of the Manuscript and Mr. Alpha Agusah and Miss Kadiatou Diallo did the design. All Authors reviewed and approved the final version.

4. Data Availability Statement: Not Applicable.

5. Research Involving Human and /or Animals: This study did not involve any humans or animals.
6. Informed Consent: Informed consent was obtained from all Participant involved in the study.

9. Acknowledgment

The authors would like to thank Professor Dr. Ramandeep Kaur, tutor of Professional Ethics and Practices and Dr. Manik Mehra, tutor of Big Data at the School of Computer Application, Lovely Professional University for their supervision, support and valuable insights during this research. Special thanks to my IoT lecturer Dr. Ashwani Kumar for his assistance with technical guidance.

10. Conclusion

The study's ramifications go beyond the short-term advantages of energy optimization. By incorporating predictive modeling into smart home systems, we open the door to more responsive and intelligent spaces that can instantly adjust to user actions and outside circumstances. This not only improves user comfort but also makes a substantial contribution to sustainability initiatives by lowering peak load demands and energy waste.

Additionally, because of its scalability, the suggested model can be used in a variety of home contexts, making it a flexible energy management tool. Combining machine learning and big data analytics will be more and more important as smart home technologies advance in order to create systems that are not only effective but also resistant to the fluctuating patterns of energy use.

Future research should concentrate on incorporating renewable energy sources like solar and wind power and broadening the dataset to include a variety of home profiles. The system's capacity to adjust and optimize in real-time may also be improved by using reinforcement learning techniques, offering a more comprehensive approach to energy management.

To sum up, this paper's developments demonstrate how predictive modeling can revolutionize smart home energy management. We can design smarter, more sustainable living spaces that complement the more general objectives of environmental preservation and energy efficiency by utilizing big data and machine learning.

11. **References**

[1] UCI Machine Learning Repository. Individual Household Electric Power Consumption Dataset. https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption

[2] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD.

[3] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation. Debnath, K. (2019). Smart Energy Management using Machine Learning Techniques. Journal of Energy Informatics.

[4] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

[5] Debnath, K. (2019). Smart energy management using machine learning techniques. *Journal of Energy Informatics*, *2*(1), 25–32. https://doi.org/10.1186/s42162-019-0072-3

[6] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[7] Li, H., & Wen, F. (2020). A hybrid ARIMA-SVM model for forecasting household electricity consumption. *Energy and Buildings*, *206*, 109566. https://doi.org/10.1016/j.enbuild.2019.109566

[8] UCI Machine Learning Repository. (n.d.). Individual household electric power consumption data set. Retrieved from https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption

[9] Zhang, Y., Wang, J., & Wang, C. (2018). Short-term electricity load forecasting using a hybrid neural network model. *Sustainability*, *10*(5), 1331. https://doi.org/10.3390/su10051331

[10] Tanwar, S., Badotra, S., Rana, A. (Eds.). (2022). Machinelearning, blockchain, and cyber security in smart environments:Application and Challenges. CRC Press.

[11] Bhutoria, A. (2022). Personalized education and artificial intelligence in the United

States, China, and India: A systematic review using a human-in-the-loop model. Computers and Education: Artificial Intelligence, 3, 100068.

[12] Hnamte V, Hussain J (2023) Dcnnbilstm: An efficient hybrid deep learning-based intrusion detection system. Telematics Inform Rep 10:10005

[13] Mohamed D, Ismael O (2023) Enhancement of an IoT hybrid intrusion detection system based on fog-to-cloud computing. J Cloud Computing 12(1):41