JOURNAL OF INNOVATION AND TECHNOLOGY
eISSN:2805-5179 Vol. 2024, No.20

Big Data and Machine Learning-Based Iot Models for Sustainable Energy Prediction

Sravani Parvathareddy[1], Vinitha Kanakambaran[2]

[1,2]EEE Department, Faculty of Engineering, New Era College of Arts Science and Technology, Botswana

[*]**Email:** sparvathareddy@neweracollege.ac.bw[1], vindira@neweracollege.ac.bw[2]

Abstract

Integrating Big Data and Internet of Things (IoT) platforms is the focus of this research, which aims to improve energy management. The problem statement is centered on the potential for development through advanced technologies and the inefficiencies in traditional energy management methods. The objectives are to analyze energy consumption patterns, develop an innovative Home Energy Management System (HEMS) architecture, and offer energy-saving solutions. Synthetic energy consumption data is generated, normalized, and divided into training and testing sets from a methodological perspective. K-nearest neighbors, Decision Trees, Support Vector Regression, and Random Forest are the machine learning models trained and evaluated. The Random Forest model outperforms other models in terms of the accuracy of its predictions of energy consumption. The integration of renewable energy sources with cutting-edge technology to revolutionize energy management practices is the essence of novelty. In conclusion, this investigation underscores the importance of utilizing advanced technologies to promote sustainable energy management, providing practitioners and policymakers with practical insights.

Keywords

Energy management, Big Data, Internet of Things (IoT), Home Energy Management System (HEMS), Machine learning, Renewable energy and Sustainability

Introduction

Smart homes are energy-efficient, green technology that reduces energy use and environmental impact. The pursuit of these benefits typically involve efficiency trading-offs, especially when green technology are too expensive or advanced to be affordable. Big Data and IoT can solve company difficulties caused by these technologies.

These technologies allow Home Energy Management System (HEMS) to enhance daily routines and user experience by controlling Electric Vehicles (EVs) and Energy Storage Systems (ESS) in houses (Aguiar et al., 2022). Automating battery-powered and model-specific appliances with convex programming frameworks reduces energy usage and user unhappiness. Smart helmets use automation and energy-saving technologies to balance energy demand and supply. Alowaidi (2022a, 2022b).

Submission: 30 September 2024; **Acceptance:** 4 November 2024

SHEMS, or smart Home Energy Management System, can reduce energy expenditure, excess power loss, and raise user happiness. However, super-automated, smart houses with IoT and Big Data establish a technical dataspace since the system can program gadgets in real time (Furstenau et al., 2022). Advanced design of interconnected infrastructures that allow immediate collection, storage, and machine learning integration in energy management optimization is an imminent action to solve the above. Fusing machine learning models like Support Vector Regression (SVR), Decision Tree Regression (DTR), Random Forest Regression (RFR), k-nearest Neighbors Regression (kNNR), and Simple Linear Regression (SLR) within energy optimization through error reduction methods like RMSE and MAE can improve energy efficiency.

Cloud computing can improve smart grid data management by supporting energy-efficient and green energy, smart grid monitoring and maintenance, location-based services (LBS), and user-aware power regulation, according to recent studies. Machine learning and other optimization methods will make smart homes energy efficient and weatherproof (Goyena et al., 2009). These new technologies have created the sustainable smart home by using renewable energy and grid-connected ESS instead of non-renewables (Aquino-Lugo, Klump, & Overbye, 2011).

This project aims to integrate renewable energy sources with Big Data and IoT platforms to improve energy management. This integration attempts to transform energy management systems to improve efficiency and sustainability.
The main objectives are:

- To analyze data using multiple machine learning methods to improve pattern recognition and energy usage estimations. Decision trees, k-nearest neighbor, support vector regression, and random forest regression are examples.
- To establish a system that processes real-time data to gain practical insights and energy-saving solutions to improve energy management and encourage energy-saving.

Literature review

Smart Home Energy Management Systems (SHEMS) are imperative for making sure the homes are using energy in an efficient way, hence meeting the needs of the renewable energy integration, energy storage, and demand response. Research reveals that energy-saving home automation is achievable through mixing Big Data, IoT, and machine learning techniques. This literature review highlights the diverse frameworks and techniques that have been suggested to improve SHEMS. Each one review below sheds light on different studies that took place in the field of energy management. Aguiar et al. (2022) suggested the IoT-enabled eco-routing decision support system, which deals with the massive amount of data to give the best routes and the lowest fuel consumption in cities. This kind of system, in case it is adjusted for the household, can be seriously pertinent to energy efficiency and the carbon footprint decrease.

Alowaidi (2022a, 2022b) also investigated IoT-based fuzzy algorithms, which are used for renewable energy management in smart homes with the main goal of optimizing the family's energy consumption through on-the-fly changes of the devices and resources by means of user interface. The inclusion of cloud computing and SHEMS together has led to this being a very widely discussed strategy of how the data can be managed at scale and the real-time monitoring and control of multiple devices can be done across the smart home environment. Chandarana & Vijayalakshmi (2014) explored cloud-centered structures to manage huge amounts of information that is produced by IoT-based smart homes. Their work highlights the need for adaptable data architectures to carry out multiple energy management features.

Aquino & Lugo (2011) suggested an agent-based smart grid control system to allow the voltage support network. Besides supporting the local voltage system in the homes, it can stabilize the network, and at the same time cutting costs, by automatically adjusting the voltage when there is a peak and off-peak hour. Cloud computing and IoT now are not the only ways to improve SHEMS. However, machine learning models have also become the key players in this field. Furstenau et al. (2022) discovered the conceptual network structure of IoT and emphasized the tool of machine learning as a vehicle of solving the residential energy demand management problem. They promoted SVR, Decision Trees, and Random Forest algorithms for predicting and optimizing the energy consumption paths based on historical data.

Bisong (2019) played an important role by way of utilizing different ML models such as the SVM and the Random Forest Regression and evaluating them by using RMSE and MAE metrics.Zeki-Suac et al. (2022) concentrated on a big data and machine learning-driven HEMS, which includes the J48 ML model and Weka API for sorting households into categories based on their energy consumption. This technique is the means to SHEMS that empowers devices to share energy across the home and thus, through smart energy management, to convert the user's excess electricity into maximum energy efficiency.The results of these studies altogether indicate that a SHEMS should have a multidimensional approach, such as IoT, big data, cloud computing, and machine learning

Methodology

a. Machine Learning and Data Preprocessing Techniques:

Data preparation and machine learning methods improve the system's prediction skills. Data analysis and prediction use regression methods including k-nearest neighbor, support vector, decision tree, and random forest. Data preparation improves energy consumption projections by ensuring data quality and integrity. Regression methods including k-nearest neighbor, support vector, decision tree, and random forest are used to predict future trends and monitor energy use. Data preparation methods are crucial for purifying and turning raw data into useful insights. This enhances energy forecast accuracy. Figure 1 represents the block diagram of a sustainable energy management system.

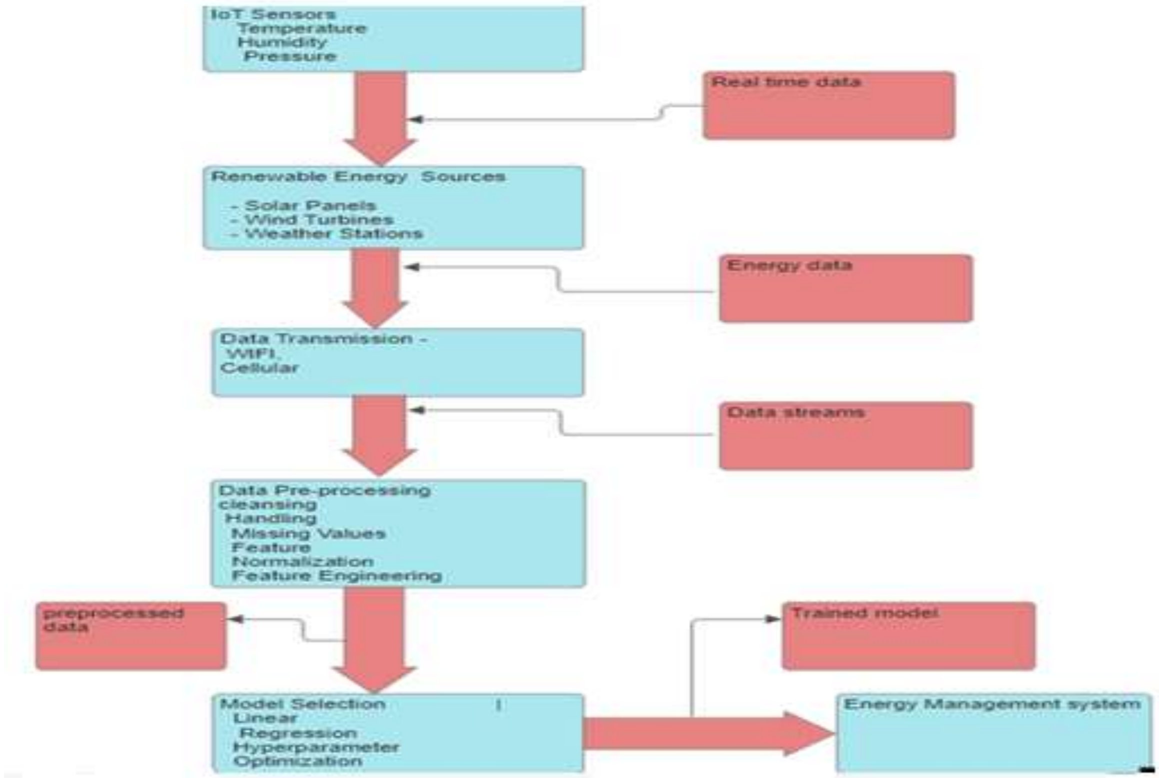

Figure 1: bBock diagram of a sustainable energy management system

b. Description of Figure 1:

Data Collection is to gather data from IoT sensors, renewable energy sources, weather stations, etc. Data Preprocessing cleanses the data, handle missing values, normalize features, etc. Model Selection is to choose appropriate machine learning algorithms for energy prediction. Model Training is used to train the selected machine learning model using the preprocessed data. Model Evaluation evaluates the performance of the trained model using appropriate performance matrices. The trained model is then deployed to make predictions on the system.

c. Integrating IoT and Big Data platform:

The system's IoT and Big Data capabilities ensure it can scale and react to evolving energy management needs to suit the performance model.

Research Process

- Energy consumption is calculated by measuring the amount of energy used over a specific period, usually in kilowatt-hours (kWh). The calculation can be determined using the formula:
  $E = P \times t$ where P represents the power consumption in kilowatts (kW) and t represents the time in hours (h).

- Mean Absolute Error (MAE) is a metric used to quantify the average magnitude of errors in a given set of predictions, regardless of their direction. The formula for Mean Absolute Error (MAE) is expressed as follows:

  MAE = $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|$ where n represents the number of observations, y_i represents the actual values, and ŷi represents the predicted values.

  Here, n represents the number of observations, y_i represents the actual value, and $\hat{y_i}$ represents the anticipated value.

Results and Discussion

Random Forest model outperformed the other models with the lowest MAE, suggesting it is better at capturing the complex relationships in the energy consumption data. The Decision Tree model also performed reasonably well but is more prone to overfitting. K-Nearest Neighbors and Support Vector Regression models showed moderate performance. This highlights the importance of using ensemble methods like Random Forest for energy management analytics to achieve more accurate predictions and better energy-saving recommendations. The results for Energy usage over time during the year is shown in Figure 2.

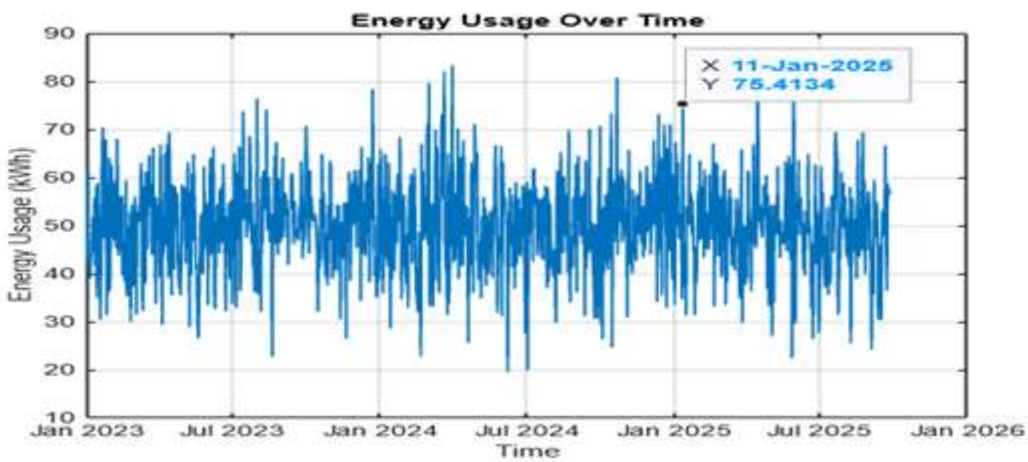

Figure 2: Energy usage over time

Model Performance comparison with Mean Absolute Error for each model is shown under Table 1 and the performance comparison between different algorithms is presented under Figure 3.

Table:1

Model	Mean Absolute Error (MAE)
Decision Tree	0.5574
K-Nearest Neighbors	0.4845
Support Vector Regression	0.4961
Random Forest	0.4297

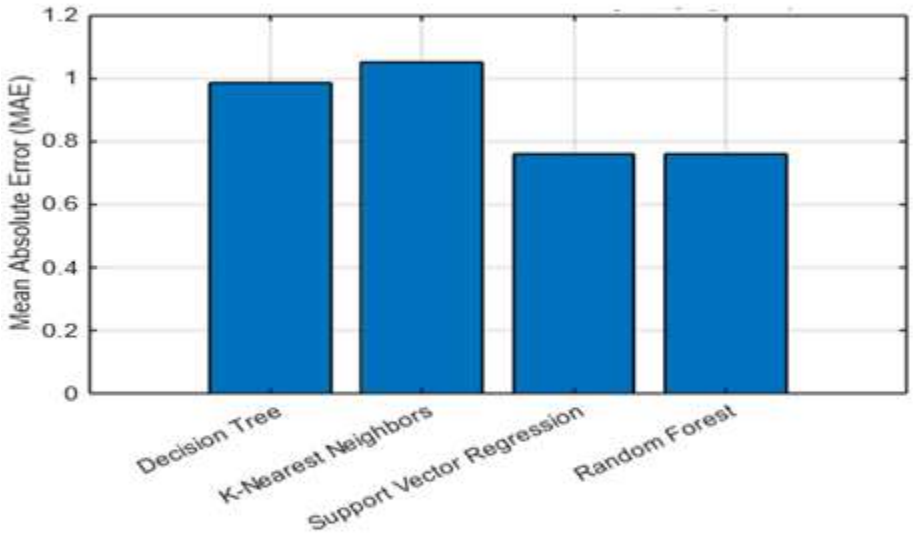

Fig: 3: Performance comparison between different algorithms

Conclusion

Integrating IoT and Big Data platforms in managing renewable energy, particularly in Home Energy Management Systems (HEMS), can revolutionize energy management strategies and sustainability programs. The recommended five-layer architecture optimizes resource utilization and energy consumption trends by applying machine learning techniques. Encouraging interdisciplinary collaborations, sponsoring research, and advocating for legislative changes can lead to a more resilient and environmentally friendly future. It is possible to modify energy management strategies in this way.

Acknowledgments

The authors would like to thank the New Era College of Arts, Science & Technology for their valuable contributions to this work.

References

Aguiar, A., Fernandes, P., Guerreiro, A. P., Tomás, R., Agnelo, J., Santos, J. L., Araújo, F., Coelho, M. C., Fonseca, C. M., D'Orey, P. M., Luís, M., & Sargento, S. (2022). MobiWise: Eco-routing decision support leveraging the Internet of Things. *Sustainable Cities and Society*, *87*, 104180. https://doi.org/10.1016/j.scs.2022.104180

Alowaidi, M. (2022a). Fuzzy efficient energy algorithm in smart home environment using Internet of Things for renewable energy resources. *Energy Reports*, *8*, 2462–2471. https://doi.org/10.1016/j.egyr.2022.01.177

Alowaidi, M. (2022b). Fuzzy efficient energy algorithm in smart home environment using Internet of Things for renewable energy resources. *Energy Reports*, *8*, 2462–2471. https://doi.org/10.1016/j.egyr.2022.01.177

Aquino-Lugo, A. A., Klump, R., & Overbye, T. J. (2011). A Control Framework for the Smart Grid for Voltage Support Using Agent-Based Technologies. *IEEE Transactions on Smart Grid*, *2*(1), 173–180. https://doi.org/10.1109/tsg.2010.2096238

Aquino-Lugo, A. A., Klump, R., & Overbye, T. J. (2011). A control framework for the smart grid for voltage support using agent-based technologies. In IEEE PES/IAS conference on sustainable alternative energy (pp. 1–7). http://dx.doi.org/10.1109/SAE.2009.5534837

Bisong, E. (2019). Google colaboratory. In Building machine learning and deep learning models on the Google Cloud platform. Berkeley, CA: A Press. http://dx.doi.org/10.1007/978-1-4842-4470-8_7

Chandarana, P., & Vijayalakshmi, M. (2014). Big data analytics frameworks. In Proc IEEE int conf circuits, syst, commun inf technol appl, Vol. 43 (pp. 0–434). http://dx.doi.org/10.1109/CSCITA.2014.6839299

Furht, B. (2011). *Multimedia Tools and Applications*. Springer.Furstenau, L. B., Rodrigues, Y. P. R., Sott, M. K., Leivas, P., Dohan, M. S., López-Robles, J. R., Cobo, M. J., Bragazzi, N. L., & Choo, K. R. (2022). Internet of things: Conceptual network structure, main challenges and future directions. *Digital Communications and Networks*, *9*(3), 677–687. https://doi.org/10.1016/j.dcan.2022.04.027

Goyena, S. G., Sádaba, Ó. A., & Acciona, S. A. (2009). Sizing and analysis of significant scale and isolated electric systems based on renewable sources with energy storage. In IEEE PES/IAS conference on sustainable alternative energy (pp. 1–7). http://dx.doi.org/10.1109/SAE.2009.5534837

He, X., Ai, Q., Qiu, C., Huang, W., Piao, L., & Liu, H. (2015). A Big Data Architecture Design for Smart Grids Based on Random Matrix Theory. *IEEE Transactions on Smart Grid*, 1. https://doi.org/10.1109/tsg.2015.2445828

Hu, H., Wen, Y., Chua, T., & Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. *IEEE Access*, *2*, 652–687. https://doi.org/10.1109/access.2014.2332453

Jena, R. (2020). An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach. *Industrial Marketing Management*, *90*, 605–616. https://doi.org/10.1016/j.indmarman.2019.12.012

Jiang, H., Wang, K., Wang, Y., Gao, M., & Zhang, Y. (2016). Energy big data: A survey. *IEEE Access*, *4,* 3844–3861. https://doi.org/10.1109/access.2016.2580581

Ju, X., Chen, V. C., Rosenberger, J. M., & Liu, F. (2021). Fast knot optimization for multivariate adaptive regression splines using hill climbing methods. *Expert Systems With Applications*, *171*, 114565. https://doi.org/10.1016/j.eswa.2021.114565

Ju, X., Liu, F., Wang, L., & Lee, W. (2019). Wind farm layout optimization based on support vector regression guided genetic algorithm with consideration of participation among landowners. *Energy Conversion and Management*, *196*, 1267–1281. https://doi.org/10.1016/j.enconman.2019.06.082

Research Anthology on Big Data Analytics, Architectures, and Applications. (2021). In *IGI Global eBooks*. https://doi.org/10.4018/978-1-6684-3662-2

Rezende, Stanley W. F., et al. (2022). "Application of Deep Learning Techniques for the Impedance-based SHM to the Oil & Gas Industry". In Jorge, Ariosto B., et al. (Eds.) Fundamental Concepts and Models for the Direct Problem, Vol. II, UnB, Brasilia, DF, Brazil, pp. 349–385. Book series in Discrete Models, Inverse Methods, & Uncertainty Modeling in Structural Integrity https://doi.org/10.4322/978-65-86503-83-8.c13

Veith, E. M. (2017). *Universal Smart Grid Agent for Distributed Power Generation Management*. Logos Verlag Berlin GmbH.

Viral, R., Tomar, A., Asija, D., Rao, U. M., & Sarwar, A. (2022). *Smart Grids for Renewable Energy Systems, Electric Vehicles and Energy Storage Systems*. CRC Press.