www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 14 Issue 11 November 2025, Page No. 28017-28020
ISSN: 2319-7242 DOI: 10.18535/ijecs/v14i11.5358

A Predictive and Simulation Framework for Detecting High Energy Consumption Periods in Smart Homes

Ghala Al-Dosari[1], Ghala Asiri[2], Ryoof Al-Qahtani[3]

[1]*CS Student*
Najran University, Najran, Saudi Arabia, 66251
[2]*CS Student*
Najran University, Najran, Saudi Arabia, 66251
[3]*CS Student*
Najran University, Najran, Saudi Arabia, 66251

Abstract
The rapid expansion of smart home technologies has enabled high-resolution monitoring of appliance-level electricity consumption. This study presents a hybrid framework that combines machine-learning-based prediction with discrete-event simulation to detect and analyze periods of elevated energy usage in smart residential environments. A Random Forest classifier was trained using appliance-level time-stamped wattage readings enriched with temporal and behavioral features. Time-series-aware cross-validation was employed to prevent data leakage and ensure realistic model performance.

To evaluate real-time applicability, the trained classifier was integrated into a SimPy-based simulation that generates synthetic appliance behavior over an extended weekly horizon. Experimental results show that the model achieves high predictive performance (accuracy ≈ 0.986) and maintains stable behavior when deployed within a dynamic simulated environment. The proposed framework provides a practical foundation for energy optimization, demand-response strategies, and intelligent automation systems in smart homes.

Keywords- ***Energy Prediction, Smart Home Analytics, Random Forest, Simulation, SimPy, High Consumption Detection***

I- Introduction

Modern smart homes collect fine-grained energy consumption data via IoT-enabled sensors, smart plugs, and automated monitoring systems [4]. Such data sources enable the development of predictive models capable of identifying periods of unusually high energy use, supporting energy optimization and intelligent device scheduling [1,4].

This research introduces a complete analytical workflow that spans data preprocessing, feature engineering, model development, and simulation integration [2,5]. The primary goal is twofold:

(1) to build a classifier that accurately predicts high consumption periods, and
(2) to embed this classifier inside a simulation environment to observe its real-time behavior under dynamic appliance activity.

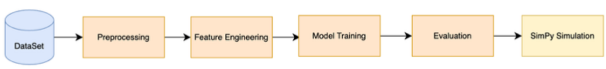

Fig. 1- demonstrates the full pipeline—from raw data ingestion to simulation-driven evaluation.

II-Dataset And Preprocessing

The dataset used in this study, titled **"Energy Consumption Dataset for Smart Homes,"** contains time-stamped wattage readings for five appliances: ventilator, PC, air conditioner (AC), lamp, and television [5].

II.I- Timestamp Processing

All timestamps were converted into Python datetime objects and sorted chronologically to preserve temporal order. This ensured correct sequencing for time-series analysis and prevented leakage between training and testing sets [2].

II.II- Feature Engineering

Several temporal and behavioral features were engineered to capture daily and weekly consumption patterns, including:

- Hour of day (0-23)
- Day of week (0-6)
- Weekend indicator (binary)
- Part-of-day segmentation

These features encode lifestyle patterns that strongly correlate with appliance activity and overall energy demand [4].

II.III- Target Construction

Total energy consumption per timestamp was computed by aggregating readings from all appliances. High-consumption periods were labeled using the 75th percentile threshold, ensuring the classifier focuses on significant demand spikes rather than normal fluctuations.

III- Model Development

III.I- Model Selection

A **Random Forest classifier** was selected due to its robustness, ability to model nonlinear relationships, and demonstrated performance on structured consumption data. The ensemble nature of the model mitigates overfitting and handles noisy appliance readings effectively.

III.II- Hyperparameter Optimization

GridSearchCV combined with TimeSeriesSplit was used for hyperparameter tuning to avoid data leakage. The optimization process explored:

- Number of estimators
- Maximum tree depth
- Minimum samples per split

This time-aware search ensures that training only uses past data to predict future observations.

III.III- Evaluation Method

A chronological 80/20 train-test split was applied. Evaluation metrics included Accuracy, Precision, Recall, F1-score, Confusion Matrix. This metric set captures both overall correctness and the classifier's ability to detect high-energy events specifically.

IV- Results And Analysis

IV.I- Confusion Matrix

Table 1- Classification Confusion Matrix

Actual \ Predicted	Low	High
Low Consumption	1078	1
High Consumption	18	264

IV.II- Performance Metrics

Table 2- Performance Metrics of the Random Forest Classifier

Metric	Value
Accuracy	0.986
Precision (Low)	0.98
Precision (High)	1.00
Recall (Low)	1.00
Recall (High)	0.94
F1 Score (Low)	0.99
F1 Score (High)	0.97

The results indicate that the model reliably identifies high-consumption periods, making it suitable for integration into simulation-driven smart home energy management systems [2,3,5].

IV.III- Feature Importance

The most influential features contributing to high-energy predictions include:

- AC usage
- Evening time periods
- Weekend activity patterns

These observations align with human activity trends.

V- Integration With Simpy Simulation

To evaluate real-time applicability, the trained classifier was embedded into a discrete-event simulation implemented using the **SimPy** toolkit [3]. The simulation generates synthetic household behavior and dynamically queries the model to predict whether each timestamp corresponds to a high-consumption period.

This integration enables:

- Predictive device scheduling
- Load balancing strategies
- Testing hypothetical what-if scenarios
- Realistic behavior modeling for smart home systems

The simulation demonstrates the feasibility of deploying the predictive model in dynamic smart home environments, where appliance usage patterns continuously change over time.

VI- Conclusion

This work presents a predictive machine-learning and simulation framework for identifying high-energy consumption periods in smart homes. The Random Forest classifier demonstrated strong predictive performance, while its integration with a SimPy-based simulation showcased the model’s capacity to operate reliably in real-time dynamic scenarios.

Future work includes incorporating occupancy detection, environmental sensor data, and sequential deep learning models such as LSTM networks to further enhance predictive accuracy and contextual awareness.

Acknowledgment

The authors would like to thank the creators of the **Energy Consumption Dataset for Smart Homes** (Arrubla-Hoyos & Severiche Maury, 2025) for providing open access to appliance-level energy data, which enabled the development and evaluation of the proposed predictive framework. Special thanks to **Ahmed Mohammed Nasser Alkheder, Assistant Professor**, for guidance and support throughout this study.

References

1. *Zeifman, M., & Roth, K. (2011). Nonintrusive appliance load monitoring: Review and outlook. IEEE Transactions on Consumer Electronics, 57(1), 76–84.*
2. *Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R News, 2(3), 18–22.*
3. *SimPy Documentation. (2022). SimPy: Discrete-event simulation for Python. Available online: https://simpy.readthedocs.io*
4. *Zhang, C.; & Li, F. (2015). Smart home energy management system using machine learning. Energy Procedia, 75, 174–179.*
5. *Arrubla-Hoyos, W.; & Severiche Maury, Z. (2025). Energy Consumption Dataset For Smart Homes [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.14768659*