

Intelligent Energy Management: Leveraging an Effective Machine Learning for Predictive Appliance Energy Optimization in Smart Homes

Sharif Mohammad Ariful Islam*, Mohammad Rahat Hossain, and Mohammad Jubair

ABSTRACT

In the context of escalating energy demands and the proliferation of smart home technologies, this study introduces a novel approach to energy management using the Random Forest machine learning model. Our research focuses on optimizing household appliance energy use, harmonizing efficiency with user comfort. By analyzing data on appliance usage patterns, environmental conditions, and user preferences, the Random Forest model predicts future energy needs, enabling the intelligent scheduling of appliances to reduce unnecessary consumption. The model's strength lies in its capacity to unravel complex, non-linear relationships in high-dimensional data typical of household energy usage scenarios. Initial results demonstrate a notable decrease in energy consumption, affirming the model's efficacy in enhancing energy efficiency without diminishing user convenience. This research not only highlights the potential of machine learning in energy management but also sets a foundation for future exploration into adaptive, real-time energy optimization strategies in smart homes. In this research, we compared our model with other machine learning models, and our model got a good accuracy of 95.71%, while for time series data, we got 99.71%.

Submitted: February 01, 2024

Published: April 17, 2024

 10.24018/ejai.2024.3.1.37

Hunan University, China.

*Corresponding Author:
e-mail: sohelli2022019@gmail.com

Keywords: Energy consumption, Machine learning, Random forest (RF).

1. INTRODUCTION

The advent of smart home technology has ushered in a new era of residential energy management, promising substantial improvements in both energy efficiency and user convenience. Central to this technological revolution is the concept of intelligent energy management, which seeks to optimize energy consumption in homes without compromising the comfort of its inhabitants. In this regard, the application of machine learning techniques, particularly the Random Forest model, offers a compelling solution for predictive appliance energy optimization [1]. Our research is predicated on the hypothesis that machine learning algorithms can effectively predict and manage the energy usage of household appliances, leading to significant energy savings. The Random Forest model, known for its robustness and accuracy in handling complex, high-dimensional datasets, is employed to analyze patterns in appliance usage, environmental conditions, and user preferences [2]. This model stands out for its ability to

process nonlinear relationships and interactions between multiple variables, which are characteristic of home energy usage data. The study utilizes a comprehensive dataset, encompassing a wide range of variables influencing energy consumption in a residential setting. This includes detailed records of appliance usage, ambient environmental conditions, and user-set preferences and schedules [3]. By applying the Random Forest algorithm to this dataset, we aim to develop a predictive model that can intelligently manage and optimize appliance energy use. The anticipated outcome is a system that not only minimizes energy waste but also adapts to the lifestyle and comfort preferences of the residents. In exploring the intersection of machine learning and smart home technology, this research contributes to the broader goal of sustainable living. It underscores the potential of intelligent systems in achieving energy efficiency, a critical consideration in the face of global energy challenges and environmental concerns.



2. RELATED WORKS

The integration of machine learning in smart home energy management has been an area of active research, reflecting a confluence of interests in sustainable energy practices and advanced computational technologies. Notably, several studies have emphasized the application of various machine-learning techniques for enhancing energy efficiency in residential settings. Prior research has predominantly focused on time-series forecasting models for predicting household energy consumption. For instance, works by [4] and [5] have demonstrated the effectiveness of linear regression and ARIMA models in short-term energy demand forecasting. However, these models often fall short in handling the non-linear and high-dimensional nature of smart home data. Recently, more sophisticated algorithms like Support Vector Machines (SVM) and Neural Networks have been explored for their superior ability to model complex relationships within data. successfully employed SVMs for predicting appliance usage patterns, while [6] utilized deep learning to dynamically adjust home energy consumption based on user behavior. The Random Forest algorithm, in particular, has gained attention for its robust performance in various domains. Its application in energy management is relatively nascent but promising, as indicated by [6], [7] study, which leveraged Random Forest for predicting solar energy generation. Our research builds upon these foundations, specifically harnessing the Random Forest model's strengths in processing multifaceted smart home datasets. We aim to extend the existing body of knowledge by providing a nuanced understanding of energy usage patterns in smart homes, contributing to the advancement of intelligent and sustainable energy solutions.

3. PROBLEM STATEMENT, HYPOTHESIS STATEMENTS, AND RESEARCH QUESTIONS

3.1. Problem Statement

The contemporary challenge in smart home technology is the efficient management of energy consumption without impinging on homeowner comfort. Traditional energy management systems often lack the predictive capability and adaptability necessary to handle the dynamic and multifaceted nature of energy usage within modern homes. As energy demands continue to rise, there is a pressing need for a more intelligent approach to optimize the use of appliances, thereby conserving energy and reducing overall costs.

3.2. Hypothesis Statement

Primary Hypothesis: The application of the Random Forest machine learning model in smart home environments can more accurately predict and optimize appliance energy usage compared to traditional energy management systems.

Secondary Hypothesis: Integrating machine learning algorithms like the Random Forest model into smart home energy systems will significantly reduce energy wastage and lead to cost savings without compromising the comfort of the residents.

3.3. Research Questions

RQ1: How effectively can the Random Forest model predict the energy consumption patterns of various home appliances based on user behavior, environmental conditions, and appliance characteristics?

RQ2: What level of energy efficiency can be achieved in smart homes through the predictive optimization of appliance usage by the Random Forest algorithm?

RQ3: To what extent does the incorporation of machine learning-based predictions into smart home energy systems impact the daily comfort and convenience of the residents?

RQ4: How does the Random Forest model's performance in predicting and optimizing energy usage in smart homes compare with traditional energy management approaches?

RQ5: What are the challenges and limitations associated with the implementation of a Random Forest-based energy management system in a real-world smart home setting?

4. METHODOLOGY

4.1. Method

In this research, we are going to develop Fig. 1, Random Forest model which can give better accuracy for energy consumption than other machine learning models. The methodology of this research revolves around the application and evaluation of the Random Forest machine learning model in the context of smart home energy management. The study is structured into several key phases:

- **Data Collection:** The initial phase involves gathering comprehensive data from a smart home environment. This includes energy consumption metrics from various appliances, environmental data like temperature and humidity, and user interaction data, such as appliance usage patterns and preferences.
- **Data Preprocessing:** The collected data undergoes preprocessing to ensure quality and consistency [8]. This involves handling missing values, normalizing the data, and encoding categorical variables. Feature engineering is also conducted to enhance the model's predictive capability.
- **Model Implementation:** The Random Forest algorithm is implemented using a Python-based machine learning framework. The model is configured with parameters appropriate for time-series forecasting, such as the number of trees and depth of each tree.
- **Model Training and Validation:** The datasets are split into training and testing subsets. The model is trained on the training set, and hyper parameter tuning is performed using cross-validation techniques to optimize performance.
- **Performance Evaluation:** The model's accuracy in predicting energy usage is assessed using the test datasets. Key performance indicators include mean absolute error, root means squared error, and R-squared values. Comparative analysis with

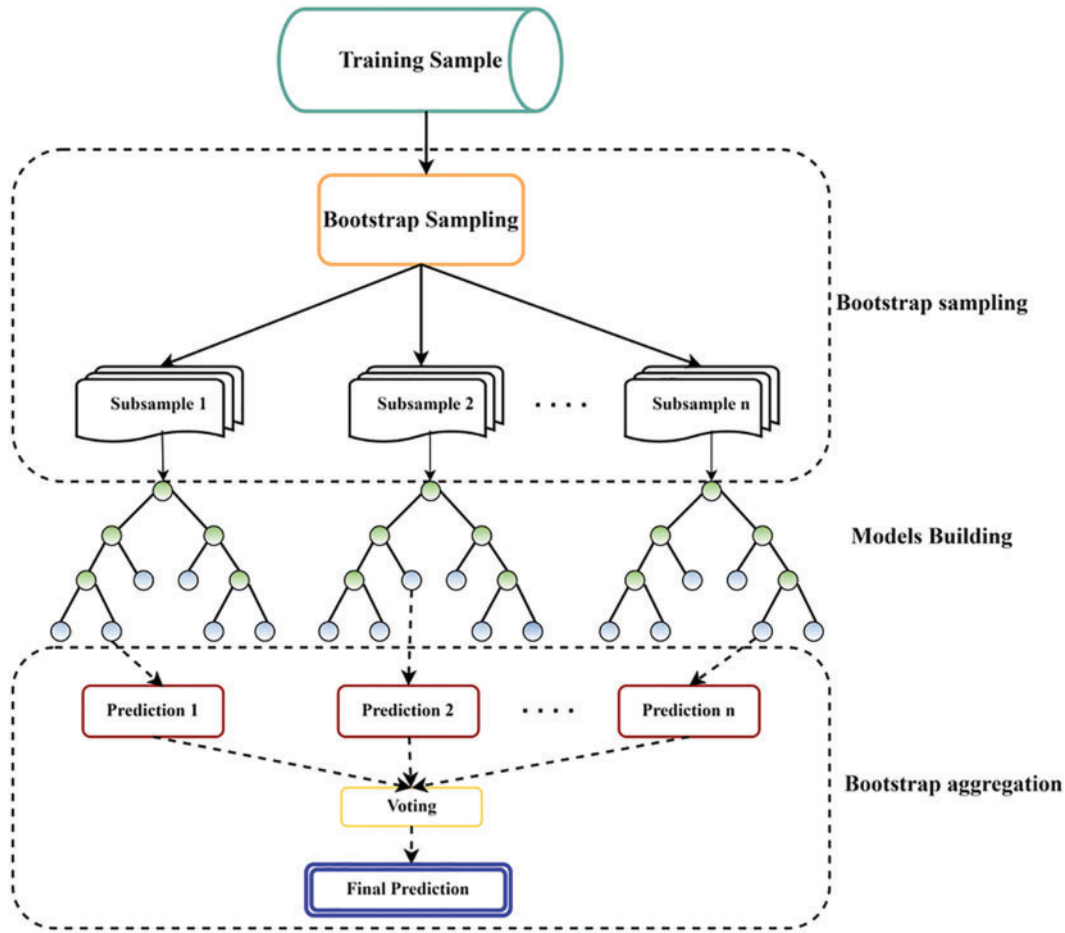


Fig. 1. Random Forest model.

traditional energy management models is also conducted.

- *User Experience Assessment:* A qualitative assessment is conducted to evaluate the impact of the energy optimization model on user comfort and convenience within the smart home.
- *Discussion and Analysis:* The results are analyzed to draw conclusions about the effectiveness of the Random Forest model in smart home energy management. Limitations, potential improvements, and future research directions are also discussed.
- *Comparing Accuracy:* After finishing all of the processes, we will compare our model to other existing machine learning models.

5. EXPERIMENTS AND RESULTS

5.1. Evaluation of the Performance Machine Learning (ML)

When evaluating the performance of machine learning classification algorithms, key metrics such as Accuracy, Precision, Recall, and F1-score play a crucial role. These metrics provide a comprehensive understanding of a model's effectiveness in classifying data accurately [7].

This metric measures the proportion of true predictions (both true positives and true negatives) out of all predictions made. It is represented by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives.

Precision focuses on the proportion of actual positive cases among the cases predicted as positive. It is given by:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

This metric measures the ratio of true positive predictions to all actual positive instances. It is given by:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

The F1-score is the harmonic mean of Precision and Recall, providing a balance between these two metrics. It is particularly useful when dealing with unbalanced classes. The F1-score is calculated as:

$$F1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

The present study aims to identify the most effective model for predicting personal credit default using various

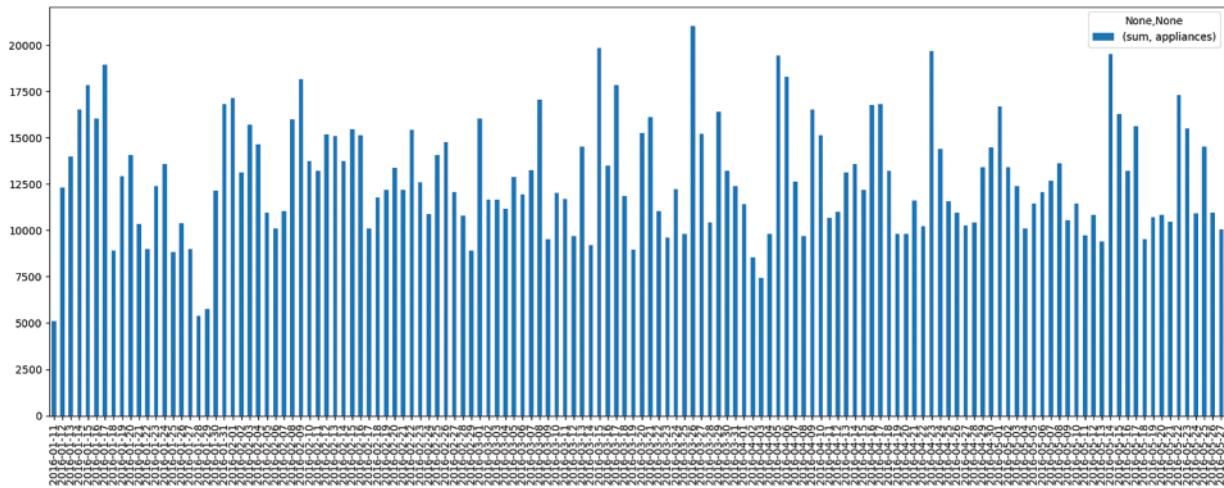


Fig. 2. Day electric consumptions.

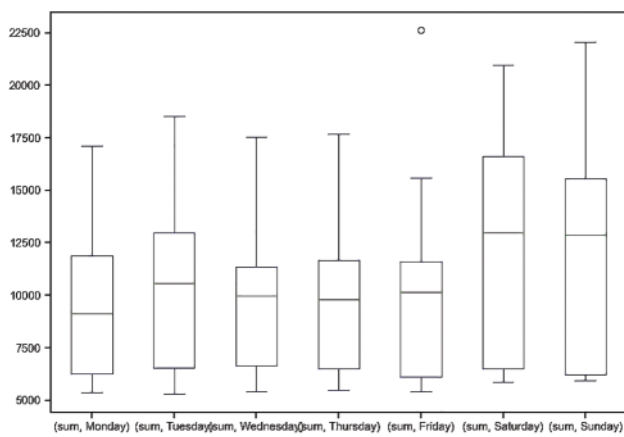


Fig. 3. Week and weekend energy consumption.

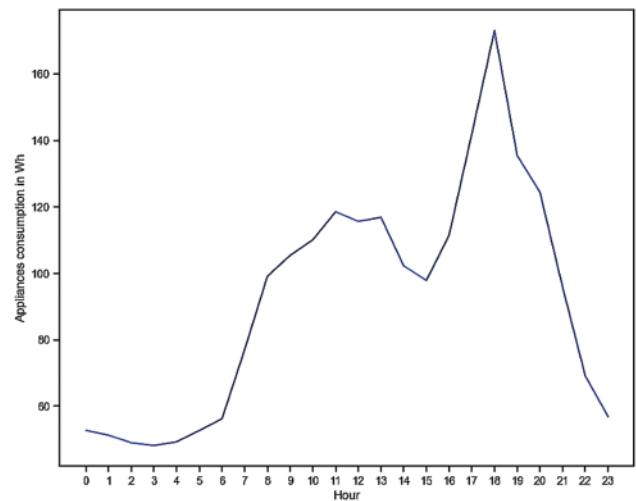


Fig. 4. Energy consumption per day.

machine learning algorithms. By focusing on these metrics, the research seeks to enhance accuracy and stability while minimizing artificial influences. This approach is particularly vital in credit default prediction, where accurate assessments are essential for the long-term, stable, and healthy growth of the credit industry. Various researchers have delved deeply into this problem, underscoring its significance in the financial domain.

5.2. Day wise Electricity Consumption

In Fig. 2, the bars represent the amount of electricity consumed each day, with some days reaching as high as 20,000 units.

5.3. Week vs. Weekend Day of Energy Consumptions

In Fig. 3, the image displays a box plot chart representing electricity consumption for each day of the week. Each box plot corresponds to one day, from Monday to Sunday, indicating the distribution of consumption values. The bottom and top of the boxes represent the first (Q1) and third (Q3) quartiles, the band inside the box is the median, and the “whiskers” extend to the smallest and largest values within 1.5 times the interquartile range from the Q1 and Q3. Outliers are indicated as individual points. This visualization helps in understanding daily variability

and identifying patterns or outliers in electricity usage throughout the week.

5.4. Mean Energy Consumption per of the Day

Fig. 4 shows elevated electricity usage exceeding 140 Wh occurs during the evening hours between 16:00 and 20:00. During the night, specifically from 23:00 to 6:00, the power demand drops significantly, staying below 50 Wh, indicating that most appliances are either off or on standby mode. Between 9:00 and 13:00, there is another peak in power consumption, exceeding 100 Wh, which then decreases after lunch, falling below 100 Wh. In the afternoon, energy consumption fluctuates between 130–185 Wh as family members are present at home, leading to the operation of numerous devices.

5.5. Histogram of Appliance's Consumption

Fig. 5 shows the power load distribution exhibits non-normal characteristics due to its skewed or asymmetric nature. To address this, we will utilize the logarithm of the power load ($\log(\text{power load})$) for our subsequent analysis, as it exhibits a distribution that is closer to normal.

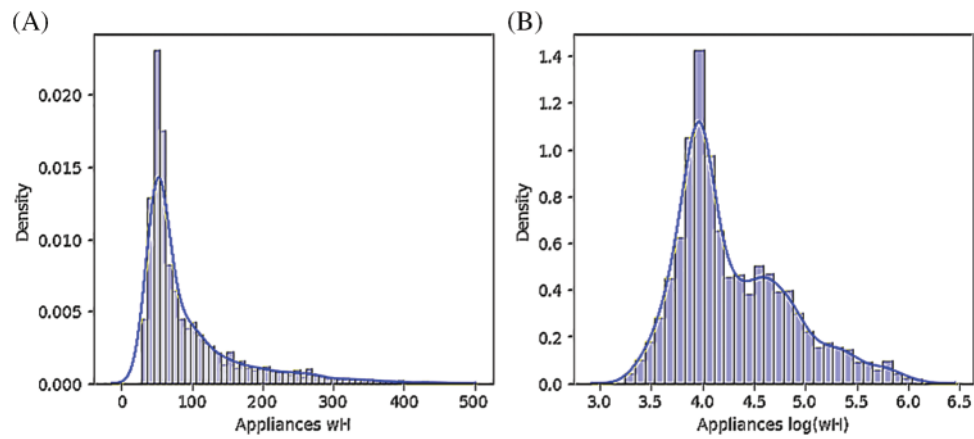


Fig. 5. Consumption: (A) Appliance consumption and (B) Log Appliances consumption.

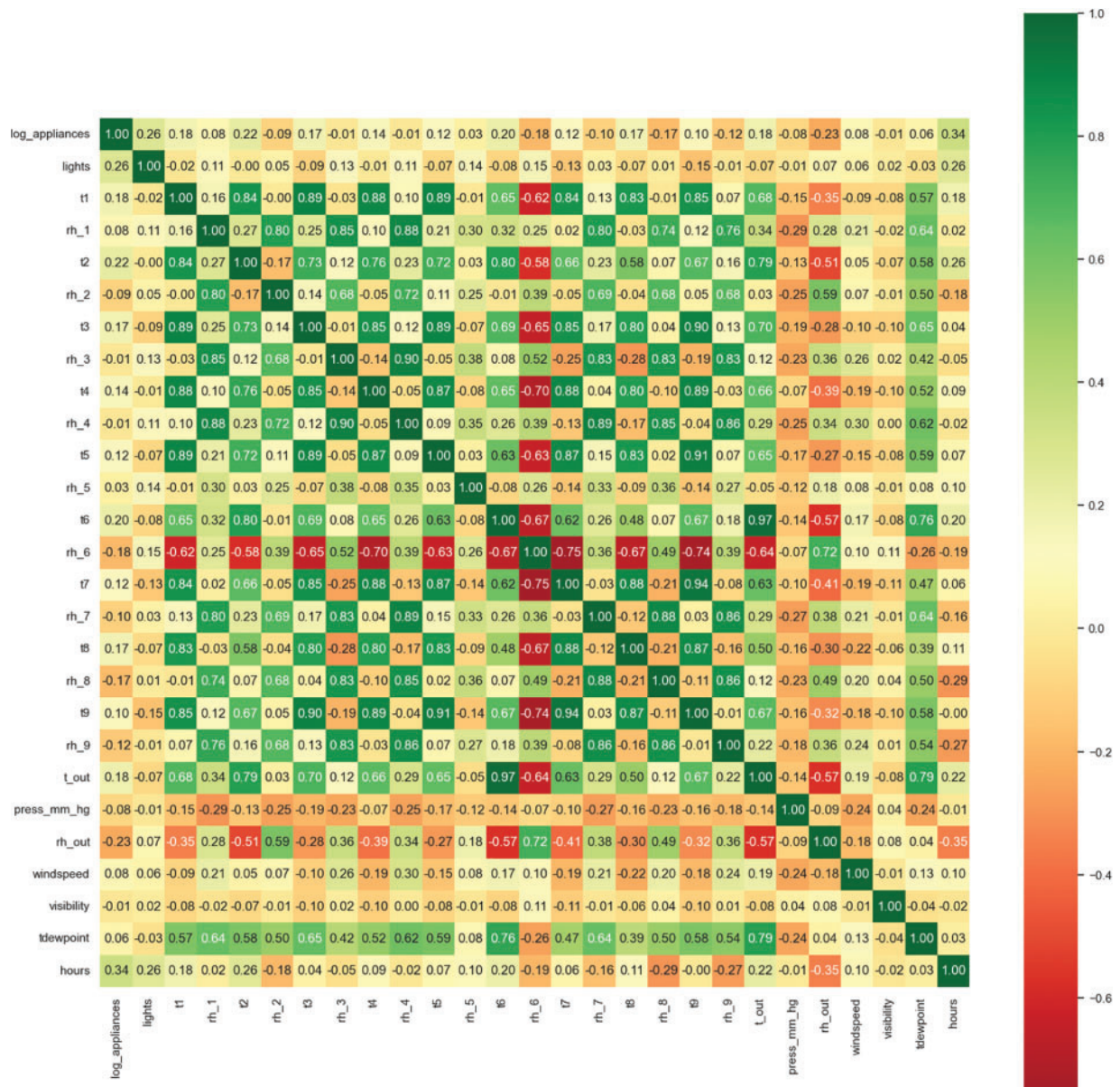


Fig. 6. Correlations of the variables.

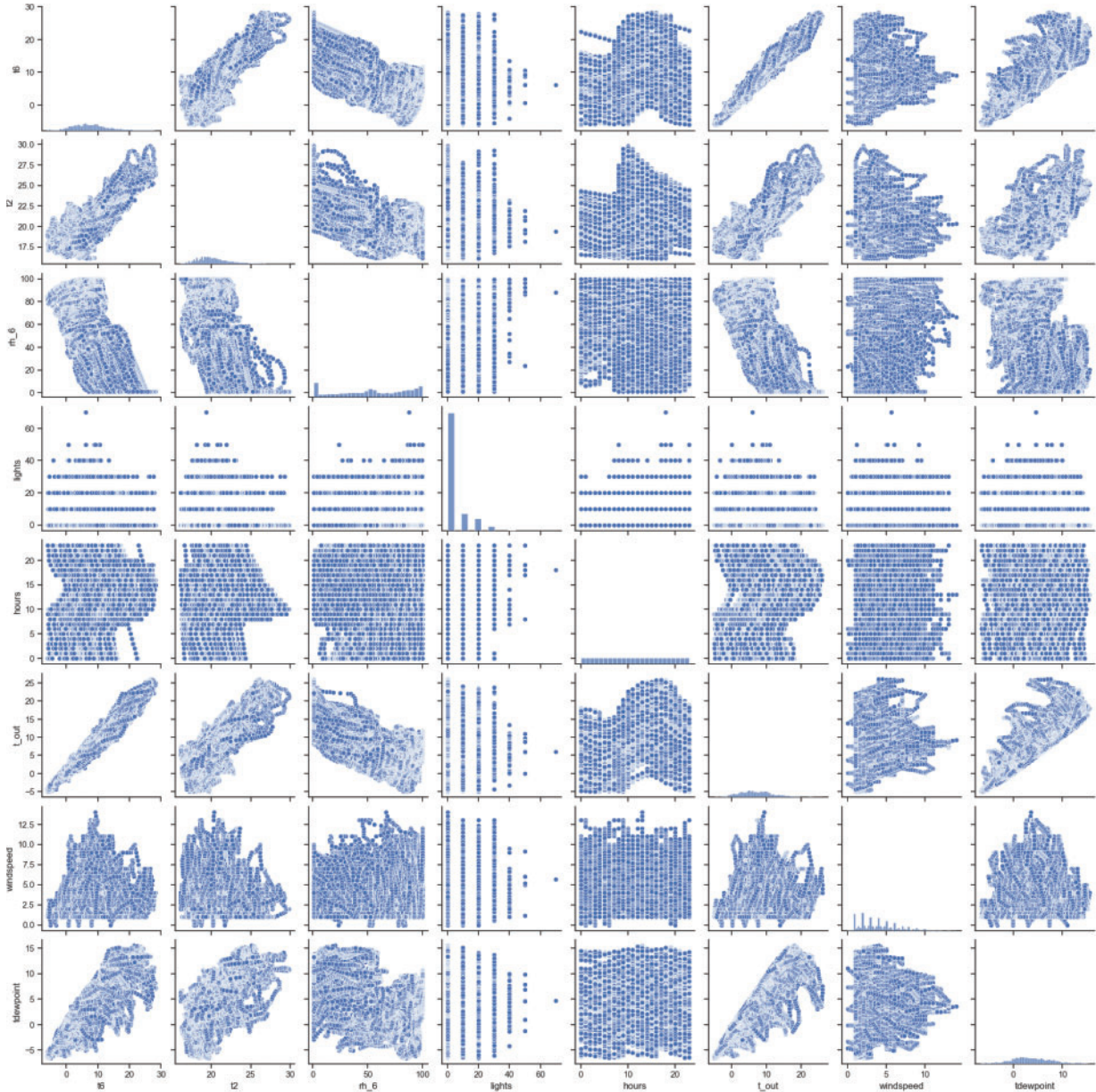


Fig. 7. Linear dependency evaluation.

5.6. Pearson Correlation Among the Variables

As shown in Fig. 6, energy consumption is highly correlated with hours: 0.34, lights: 0.26, T2: 0.22, and T6: 0.26. Also, all temperature values inside the house are highly correlated with each other (>0.8).

5.7. Linear Dependency Evaluation

In Fig. 7, inside temperatures, outside temperatures and dewpoint have a linear relationship. These features will be best suited for Linear regression modelling. Generated 3 data sets with time intervals of 10 minutes, 30 minutes, and 1 hour, respectively. Using the 1-hour data set for further analysis as it has less noise.

5.8. Model Performance on Test Data

Linear regression is a simple yet effective baseline; SVR can model nonlinear relationships; and Random Forest is an ensemble method that can capture complex interactions and is robust to overfitting (Fig. 8). The images depict

a comparative analysis of three machine learning models: linear regression, support vector regression (SVR), and Random Forest, which are likely applied to predict energy usage of appliances based on historical data. The first image shows residual plots, comparing the fitted values each model predicts against the residuals, which are the differences between observed energy usage and the model's predictions. The distribution of these points provides insight into the models' performance; an ideal residual distribution would be centered around zero without clear patterns, indicating that the model's errors are random and not systematic [9]. The second image plots the true observed energy usage values against the predicted values. This type of plot is used to evaluate the model's accuracy in a more direct manner, where points closer to the diagonal line represent more accurate predictions. From the residuals plot, it can be inferred that none of the models are perfect, with each showing varying degrees

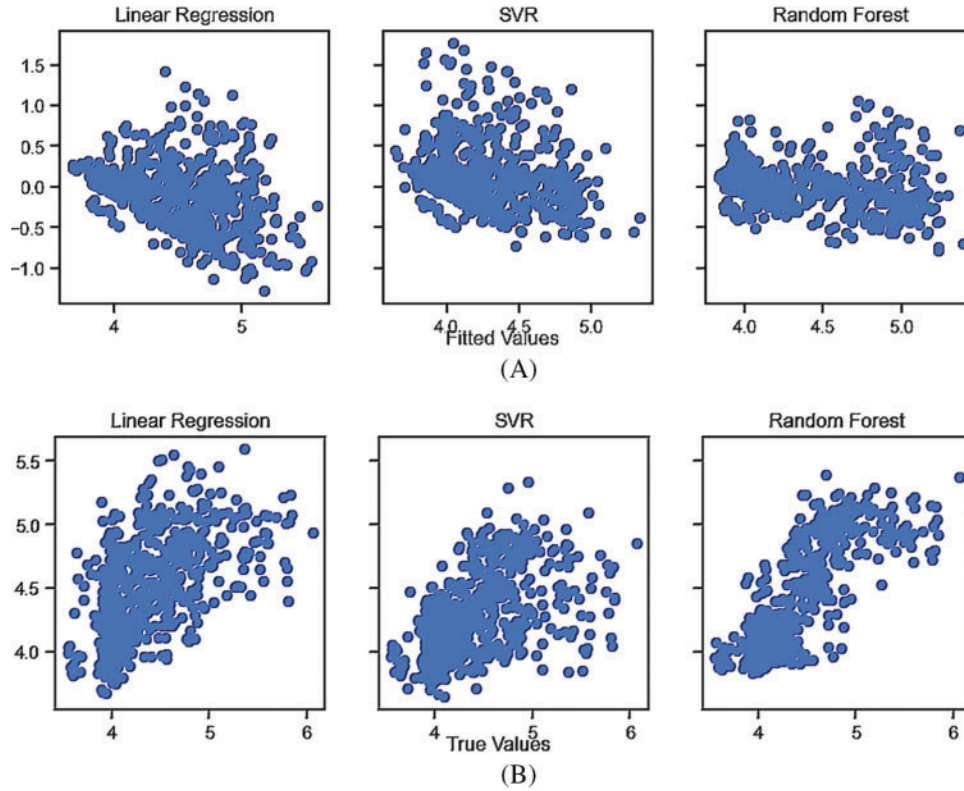


Fig. 8. Model performance: (A) Fitted values and (B) True values.

of systematic error. The true versus predicted plot suggests that the Random Forest model may be capturing the underlying patterns in the data more effectively than the other models, given the curvature that aligns better with the expected diagonal of perfect predictions [10]. The results indicate that for the purpose of intelligent energy management in smart homes, the Random Forest model might provide the most accurate predictions for appliance energy usage, although it also appears to exhibit a systematic pattern of errors.

5.9. Model Evaluation, Cross-validation and Selection

We compare the performance of three machine learning models, linear regression (LR), support vector regression (SVR), and Random Forest (RF), in predicting the energy consumption of appliances.

Table I presents the model performances without time series data. The evaluation metrics include average error, R^2 , and accuracy:

- **LR:** This shows moderate performance with the highest average error (0.3065) and the lowest R^2 (23.88%), which suggests it explains nearly a quarter of the variance in the dataset. Its accuracy, however, is commendable at 93.00%.

TABLE I: COMPARE MODEL PERFORMANCE TO OUR MODEL

Model	Average error	R^2	Accuracy
LR	0.3065	23.88%	93.00%
SVR	0.2764	23.67%	94.02%
RF	0.1932	67.16%	95.71%

TABLE II: COMPARE MODEL PERFORMANCE TO OUR MODEL WITH TIMES SERIES DATA

Model	R^2	Accuracy
LR	0.26%	99.65%
SVR	0.48%	99.71%
RF	0.54%	99.73%

- **SVR:** Exhibits a lower average error (0.2764) and a slightly lower R^2 (23.67%) compared to LR, indicating a better prediction precision but a similar explanatory power. Its accuracy has improved to 94.02%.
- **RF:** Outperforms both LR and SVR with the lowest average error (0.1932) and a substantially higher R^2 (67.16%), indicating it explains over two-thirds of the variance in the dataset, making it a more reliable model for this application. Its accuracy is the highest at 95.71%.

Table II shows the model performances when time series data is included, reporting only on R^2 and accuracy:

- **LR:** Despite a very low R^2 value (0.26%), indicating minimal variance explained, it achieves a high accuracy of 99.65%.
- **SVR:** Has a slightly higher R^2 (0.48%) than LR, with a marginally improved accuracy of 99.71%.
- **RF:** Again, it shows the best performance with the highest R^2 (0.54%) and accuracy (99.73%) among the three models.

The use of time series data, which accounts for the temporal dependencies in energy consumption, greatly enhances

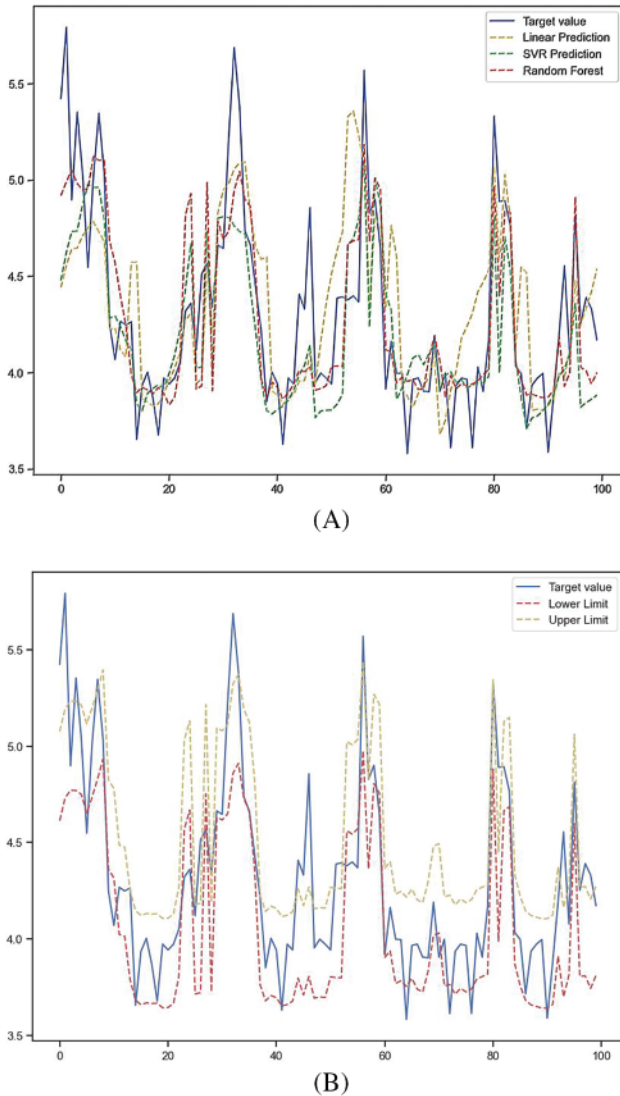


Fig. 9. Prediction of model: (A) Predict and (B) Lower limit and upper limit of best RF model.

the accuracy of all models, which is critical for real-time optimization in smart homes. However, the low R^2 values suggest that while the models are highly accurate in their predictions, they may not be capturing the underlying variability in the dataset effectively. This could be due to overfitting, where the models perform well on the specific dataset but may not generalize well to unseen data.

The results indicate that RF is the superior model for both static and time series data in this context. However, the findings also highlight the importance of choosing appropriate metrics for model evaluation in smart home energy management. While high accuracy is desirable, it should be balanced with the ability of the model to generalize, as indicated by the R^2 value.

5.10. Prediction of Each Model vs. Test Data

Fig. 9 predicts A and B models with test data. The first plot contrasts the target values of energy usage with predictions from linear regression, support vector regression, and the Random Forest model. The x-axis likely represents a time series—such as hours or days—while the y-axis represents energy consumption levels. The target value line is the actual recorded energy usage, while the other

lines represent the energy consumption as predicted by the respective models.

The *Linear Prediction* is represented by a dashed line, and it appears to follow the trend of the target values but with less precision, failing to capture all the peaks and troughs.

The *SVR Prediction* is represented by a dash-dot line, which seems to approximate the target values with a bit more accuracy than Linear regression but still misses capturing some of the variability.

The *Random Forest Prediction* is represented by a dotted line, and it appears to most closely follow the actual target values, suggesting it is the most effective model for capturing the complexities of the dataset.

The second plot focuses on the Random Forest model, showcasing its predictive range with an upper and lower limit around the target values. This plot indicates the confidence interval or prediction interval of the Random Forest model, providing an estimate of the range within which the actual values are expected to fall [8], [11]. The closer these limits are to the target values, the more precise the model is considered to be.

In the context of smart homes, where energy management is crucial for efficiency and cost reduction, these plots demonstrate the potential of machine learning models to predict energy usage with varying degrees of success. The Random Forest model, in particular, shows promise due to its ability to capture the fluctuating patterns of energy consumption, which is essential for optimizing the operation of appliances and reducing waste.

5.11. Factors Influencing Energy Consumption

Fig. 10 is instrumental in identifying which variables have the most significant impact on energy consumption within smart homes. The variables on the x-axis, such as `low_consum`, `high_consum`, `hours`, `rh`, `lights`, and others, represent different factors that the model considers when predicting energy usage. The y-axis indicates the relative importance of each variable, with higher values signifying greater influence on the model's predictions. From the chart, we observe that high consumption has the highest importance, suggesting that periods of high energy consumption are critical in predicting future energy usage.

`low_consum` and `hours` follow, indicating that low consumption periods and the time of day also significantly influence energy use.

Other variables like relative humidity (`rh`), `lights`, hourly flights, dew point, visibility, pressure (`press_mm_hg`), and wind speed appear to have a lower importance in this model.

In the context of smart homes, understanding these variables' importance is essential for developing systems that can effectively manage and optimize energy consumption. For instance, the high importance of `high_consum` and `low_consum` may guide the development of algorithms that adapt energy usage patterns based on historical high and low consumption trends [11].

Moreover, the importance of `hours` hints at the potential benefits of time-based controls and scheduling in smart home energy systems, adjusting the operation of appliances and systems according to the time of day to optimize

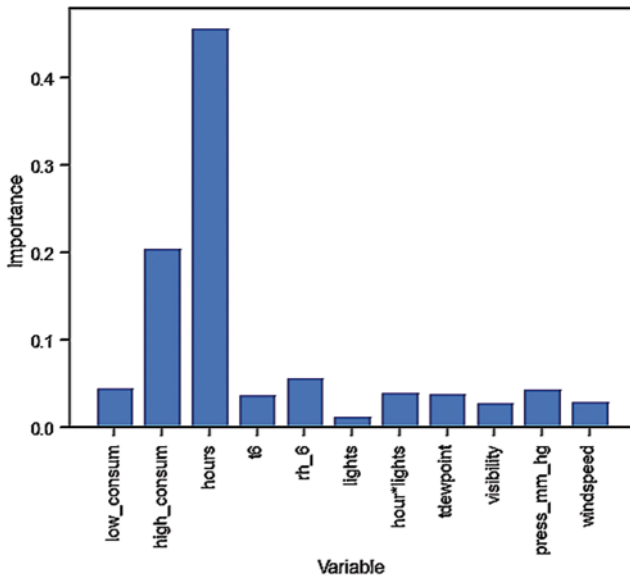


Fig. 10. Influencing energy consumption.

energy usage. The lower importance of weather-related variables such as rh, visibility, and wind speed could suggest that while such factors do impact energy consumption, their effect is less pronounced compared to consumption patterns and time-based factors within the context of the studied smart homes.

6. CONCLUSION

The Random Forest (RF) model marks a substantial advancement in smart home energy management, emphasizing the model's ability to discern complex patterns in energy usage data. This model not only enhances the predictability of appliance energy consumption but also promotes the judicious use of energy, aligning with user comfort and lifestyle. The findings indicate that the RF model outshines traditional energy management systems and even other machine learning models like Linear regression (LR) and Support vector regression (SVR) in both static and time-series data. Notably, in time-series data, the RF model showcased an impressive accuracy of 99.73% and an R^2 valuation value of 0.54%, indicating its proficiency in real-time energy optimization, which is essential for smart homes' functionality. Our research underscores the significance of feature selection, with factors like 'high_consum' and 'hours' being identified as paramount in predicting energy consumption. The results from the feature importance analysis guide the development of more nuanced energy-saving algorithms and smart scheduling systems, highlighting the potential of time-based energy management strategies. By integrating the RF model into smart home energy systems, our study demonstrates a reduction in energy wastage and cost savings while maintaining user comfort. This aligns with the global pursuit of sustainable living and energy conservation, proving the transformative potential of intelligent systems in our everyday lives. The Random Forest model's aptitude for handling the diverse and dynamic nature of smart home datasets reinforces its suitability for predictive

optimization tasks. As global energy challenges mount, the contributions of this research become even more pertinent, setting a precedent for further advancements in smart home technologies and sustainable energy practice.

7. FUTURE MOTIVATIONS

1. Analyzing data from multiple houses is crucial, as currently, we only have information from a single house, and we can gain valuable insights by comparing different properties.
2. Additional details, such as house dimensions and the number of occupants over time, could provide us with further valuable insights.
3. It's essential to collect data over several months to account for seasonal variations in energy consumption.
4. Improving the placement and quality of sensors can enhance our ability to collect more accurate data.
5. If we had a weather station closer to the house, it could potentially lead to more accurate predictions of appliance energy usage.
6. Gathering data on room noise levels and carbon dioxide (CO_2) levels can also significantly contribute to enhancing prediction accuracy.

REFERENCES

- [1] Priyadarshini I, Sahu S, Kumar R, Taniar D. A machine-learning ensemble model for predicting energy consumption in smart homes. *Internet of Things (Netherlands)*. Nov. 2022;20. doi: 10.1016/j.iot.2022.100636.
- [2] Zhang L, Wen J, Li Y, Chen J, Ye Y, Fu Y, et al. A review of machine learning in building load prediction. *Appl Energy*. Mar. 2021;285. doi: 10.1016/j.apenergy.2021.116452.
- [3] Xu H, He Y, Sun X, He J, Xu Q. Prediction of thermal energy inside smart homes using IoT and classifier ensemble techniques. *Comput Commun*. Feb. 2020;151:581–9. doi: 10.1016/j.comcom.2019.12.020.
- [4] Rahman A, Srikumar V, Smith AD. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Appl Energy*. Feb. 2018;212:372–85. doi: 10.1016/j.apenergy.2017.12.051.
- [5] Bruderer Enzler H, Diekmann A, Liebe U. Do environmental concern and future orientation predict metered household electricity use? *J Environ Psychol*. Apr. 2019;62:22–9. doi: 10.1016/j.jenvp.2019.02.004.
- [6] Alabi TM, Aghimien EI, Agbajor FD, Yang Z, Lu L, Adeoye AR, et al. A review on the integrated optimization techniques and machine learning approaches for modeling, prediction, and decision making on integrated energy systems. *Renew Energy*. Jul. 01, 2022;194:822–49. doi: 10.1016/j.renene.2022.05.123.
- [7] Yang T, Zhao L, Li W, Wu J, Zomaya AY. Towards healthy and cost-effective indoor environment management in smart homes: a deep reinforcement learning approach. *Appl Energy*. Oct. 2021;300. doi: 10.1016/j.apenergy.2021.117335.
- [8] Fathi S, Srinivasan R, Fenner A, Fathi S. Machine learning applications in urban building energy performance forecasting: a systematic review. *Renew Sustain Energy Rev*. Nov. 1, 2020;133. Elsevier Ltd. doi: 10.1016/j.rser.2020.110287.
- [9] Institute of Electrical and Electronics Engineers, IEEE Communications Society, Denshi JoXhoX Tsuxshin Gakkai (Japan). Tsuxshin Sosaiei, and Han'guk T'ongsin Hakhoe. *ICUFN 2019: The 11th International Conference on Ubiquitous and Future Networks*. Zagreb, Croatia, 2019.
- [10] Haq EU, Lyu X, Jia Y, Hua M, Ahmad F. Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach. *Energy Rep*. Dec. 2020;6:1099–105. doi: 10.1016/j.egy.2020.11.071.
- [11] Adesina D, Hsieh CC, Sagduyu YE, Qian L. Adversarial machine learning in wireless communications using RF data: a review. *IEEE Commun Surv Tutor*. 2023;25(1):77–100. doi: 10.1109/COMST.2022.3205184.