Energy Efficiency Optimization for Smart Homes

Kadir Gökdeniz
Dept. of Computer Engineering
Ankara University
Ankara, Turkey
20290344@ogrenci.ankara.edu.tr

Mehmet Bayram Alpay
Dept. of Computer Engineering
Ankara University
Ankara, Turkey
20290310@ogrenci.ankara.edu.tr

Ertuğrul Sert
Dept. of Computer Engineering
Ankara University
Ankara, Turkey
21290772@ogrenci.ankara.edu.tr

***Abstract*— This study offers a thorough strategy for improving smart home energy efficiency using modern techniques for data mining. Through the utilization of predictive modeling and real-time data collection from smart home devices, such as sensors and appliances, the system is capable to anticipate user behaviors and consumption patterns, therefore enabling proactive energy management. A customized and effective balance between user enjoyment and energy saving is ensured by the adaptive energy optimization model, which is driven by machine learning algorithms and constantly improves plans based on historical data and user input. the method has been proven to be successful through extensive simulations and real-world testing, which show considerable energy saving by addressing the dynamic nature of smart home settings and ensuring device compatibility, the solution helps to shape an intelligent and sustainable energy management for the future. We have with categorizing smart home technologies such as flexible, semi-flexible, non-flexible, we implanted advanced data mining techniques such as linear regression, random forest regression, support vector machine, LTSM and XGBoost. In the study paper achieved significant energy efficiency for smart homes with data mining.**

Keywords—Smart Homes, Energy Efficiency, Artificial Intelligence, Machine Learning, Clustering, Classification, Optimization, Linear Regression, XGBoost, Random Forest Regression, Data Mining, Support Vector Machine, LTSM

I. Introduction

It is expected that smart houses would use a significant amount of energy due to the increasing number of smart home appliances. This is a result of the introduction of a number of new household products that improve customers' quality of life [1]. In actuality, this comfort covers the expense of the smart home appliances required to support contemporary, rapidly evolving technology [2]. Among these are smart robots, microwaves, refrigerators, and dishwashers [3].

Energy-related problems and climate change have forced the energy system to adapt during the last few decades [4]. This dilemma will get harder as the demand for energy rises in the ensuing decades [5]. The globe is facing a significant challengein energy efficiency as a result of the rising demand for energy. But this can be challenging because of how complicated the processes and systems are. Predicting energy efficiency with data is one possible solution to this issue. Data mining may be used to address many different kinds of inefficiencies.

The goal of this project is to analyze the complexity of energy usage in smart homes in order to offer customized optimization tactics with data mining. We set out to construct sustainable, effective and smart places to live that not only satisfy the needs of the present but also anticipate the challenges of the future via a set of flexible, semi-flexible, and non-flexible categories.

In this project, we will commence by providing an overview of the relevant studies conducted in the field, followed by an exposition of our perspectives on the existing systems documented in the literature. Subsequently, we will elucidate the procedures employed in preprocessing the data, expounding on the methodologies utilized alongside a presentation of their theoretical underpinnings. Additionally, we will delve into the intricate details of the error metrics chosen for evaluation, offering comprehensive insights into their functionalities and applicability. The ensuing section will encompass a discussion of the shortcomings identified within the literature review, articulating how this study aims to address these gaps, culminating in the presentation and analysis of the outcomes. Finally, we will encapsulate the entirety of this research undertaking through a concise summary, encapsulating the core essence and contributions of the study

II. Related works

Smart home appliances' increasing numbers forecast significant energy consumption in smart homes. This trend has emerged due to the introduction of numerous new home appliances that aid users in leading a better life [6]. Indeed, the cost of smart home devices required to support modern and continually evolving technologies is covered by this convenience [7].

Dishwashers, refrigerators, microwaves, and smart cars are just a few examples [8].

In the last decade, energy systems have had to evolve due to energy issues and climate change [9]. With an anticipated increase in energy demand in the next decade, this challenge will become more arduous [10]. The rise in energy demand has turned energy efficiency into a major global challenge. However, due to the complexity of systems and procedures, addressing this challenge can be difficult. Using data to predict energy efficiency presents a potential solution. Many types of inefficiencies can be resolved using data.

Machine learning-based smart home services have garnered significant attention, contributing to the understanding and implementation of smart living spaces through various research endeavors. One study focused on characterizing low-voltage electricity customers, employing clustering and classification techniques to analyze historical data [11]. The research underscores the significance of clustering algorithms such as K-medoids and K-means in determining consumption patterns and customer classifications [12]. In the domain of home energy management, a new method utilizing deep learning (DRL) was introduced to optimize photovoltaic self-consumption and enhance user comfort and energy savings [13]. Investigating energy consumption patterns and interconnections among devices in a smart city context using data mining techniques contributes to a better understanding of consumer behavior and preferences [14]. This study, based on Moroccan energy consumption data, provides valuable insights for the development of energy management systems. Machine learning techniques play a central role in optimizing energy usage in smart homes [15]. The study assesses various methods, including deep learning, reinforcement learning, and decision trees, demonstrating their effectiveness in identifying models and forecasting future consumption [16]. Developing models predicting household appliance energy consumption involves the use of long and short-term memory networks, highlighting the benefits of deep learning in energy consumption prediction [17]. Another article explored the implementation of machine learning algorithms in a smart home context, employing decision tree regression, random forest regression, additive tree regression, gradient boosting, historical gradient regression, and deep neural networks for data analysis and prediction [18]. Lastly, a study focused on data mining and analysis to forecast electrical energy consumption, classifying electricity consumers, and employing decision trees, random forests, Naïve Bayes classifiers, and SVM algorithms for efficient data analysis and prediction [19].

Home Energy Management Systems (HEMS) and Demand Response (DR) programs incentivize residential users to reduce their electricity consumption. These programs define consumption patterns that differ from typical habits due to factors such as changing electricity costs or incentive schemes. These initiatives are widespread in Europe and the United States, aiming to regulate timing, total power consumption, and immediate demand.

These programs encourage consumers to reduce electricity usage during peak hours by offering financial incentives or discounts. Reward-focused programs offer discounted rates or bill credits to participants, utilizing methods like direct load control or curtailment programs. Price-based programs, meanwhile, use various pricing structures during different time frames to help consumers optimize energy usage.

HEMS manages household appliances using smart technologies, optimizing power consumption and scheduling energy use during critical periods in response to demand signals. These systems employ smart control algorithms like rule-based controllers, artificial neural networks (ANNs), fuzzy logic control (FLC), and adaptive neural fuzzy inference systems (ANFIS).

Rule-based controllers define behavior through conditional rules and are used in areas such as load distribution, device priority settings, and load shifting strategies. FLC optimizes programs based on weather forecasts and electricity prices using fuzzification and defuzzification stages, while ANFIS plans energy with multi-layer feedback mechanisms. Each control technique has its advantages and limitations; FLC requires appropriate rules, while ANFIS needs extensive data and training. Comparatively, ANN excels in learning complex functions effectively.

These systems provide cost savings and reduced energy consumption to consumers. HEMS utilizes smart technologies and different control algorithms, yet ANN proves more effective in learning complex functions and offers real-time performance. Figure 1 shows the advantages and consequences of ANFIS, ANN, and FLC results.

Figure 1

ANN Controller	FLC Controller	ANFIS Controller
1. Mathematical model is not required	1. Mathematical model is not required	1. Mathematical model is not required
2. Complex design and implementation	2. Easy design and implementation	2. Moderately complex design and implementation
3. Normal structure	3. Simple structure	3. Complex structure
4. Can achieve good performance if appropriate activation function, training data, and number of nodes are selected	4. Can achieve good performance if proper parameters in the rule-based algorithm and type of membership functions are selected	4. Can achieve good performance if suitable training data and type of membership function are selected
5. Requires learning process when designing the controller	5. Requires no learning process when designing the controller	5. Requires learning process when designing the controller

In the literature review, current works concentrate on three distinct systems, delineating their respective pros and cons. It is recognized that constructing ANFIS (Adaptive Neuro-Fuzzy Inference Systems) and ANN Fuzzy systems (Neuro-Fuzzy Controllers) is intricate. The process involved in building these systems is notably complex. On the other hand, developing FLC (Fuzzy Logic Controllers) systems is comparatively simpler in the construction phase compared to the former two. Nevertheless, acquiring genuine fuzzy clusters and adaptable rules for FLC systems necessitates considerable expertise.

III. DATA PREPROCESSING

Pre-processing is one of the most important stages for any project. That's why we applied different pre-processing processes for each dataset while writing. The methods we applied changed depending on the method used because we decided to apply more advanced operations in the XGBoost model. Other than that, the transactions were generally similar or the same. We explain each process here below.

Resampling: This method is involved in changing or editing the number of observations in the dataset. It can generally be preferred for problems such as class imbalance, but instead of making second-by-second predictions, we made predictions over longer periods of time by resampling the project with separate 10-minute, 30-minute, and 60-minute periods.

Normalization: Normalization is a data preparation technique applied in artificial intelligence to bring data to the same scale. This process increases the performance of the model and simplifies the training process. Normalization also helps make it easier to compare the results of metrics at the end of training. Especially since we did the resampling process on own dataset, values were very high, and thanks to the Min-Max Scaler, the model adapted to this process more easily.

Data Splitting: It is a very important task to divide the dataset into 3 different groups train, validation, and test in pre-processing processes. The model is trained with the train set, the parameters are adjusted with the validation set and it can be understood how generalizable the model is, and the results of the model are tested with the test set. This process is one of the most common process we see in projects, and in the project the train set is generally the majority.

Feature Creation: It is a very important method used in machine learning projects and data analysis. The purpose of this process is to derive new features by using the features in the existing dataset. This process can generally be used for needs such as helping the model learn better, completing missing information, and reducing and increasing complexity. While writing code, we looked at the feature importance values in order to notice better features and accordingly, we removed the ones with lower values to avoid overfitting.

```python
def create_features(df):
  """Time Series Features"""
  df = df.copy()
  if 'timestamp' in df.columns:
      df['hour'] = df['timestamp'].dt.hour
      df['hourly_mean'] = df.groupby('hour')['power'].transform('mean')
      df['hourly_min'] = df.groupby('hour')['power'].transform('min')
      df['hourly_max'] = df.groupby('hour')['power'].transform('max')
      df['dayofweek'] = df['timestamp'].dt.dayofweek
      df['weekday'] = (df['dayofweek'] < 5).astype(int)

      # Lag özellikleri
      df['lag_1'] = df['power'].shift(1)

      # Diff özellikleri
      df['diff_1'] = df['power'].diff(1)
      df['diff_2'] = df['power'].diff(2)
  return df
```

IV. DATA MINING TECHNIQUES

In the study, machine learning methods, such as Random Forest Regression, Support Vector Machines, Linear Regression, Long Short-Term Memory (LSTM) networks, and Extreme Gradient Boosting (XGBoost), used in investigation. This broad range of techniques made it possible to examine the dataset in depth while utilizing the advantages of each approach for dealing with particular data structures and patterns. Sequential dependencies were well captured by LSTM networks, and intricate non-linear interactions were superbly modeled by Random Forest Regression. Support Vector Machines demonstrated efficacy in classification tasks, especially where there were clear class distinctions. To complement XGBoost's ensemble-based approach, Linear Regression offered interpretability and simplicity in comprehending basic data relationships.

*a) **Long Short-Term Memory (LTSM)***

Recurrent neural networks (RNNs) of the long-term dependency type (LSTM) are used to identify long-term dependencies in sequential data. In contrast to conventional RNNs, LSTMs use gating mechanisms (input, forget, and output gates) in conjunction with memory cells to obtain and update data in a selective manner throughout a number of time steps.

Mathematical Formulation

The Key equations LTSM are:

- **Input Gate:** $i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi})$
- **Forget Gate:** $f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf})$
- **Cell State:** $g_t = \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg})$
- **Output Gate:** $o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho})$
- **Hidden State:** $h_t = o_t \cdot \tanh(C_t)$

-*Xt is the input at time step*

-*ht is the hidden state at time step*

-C_t is th ecell state at time step

-σ *is the sigmoid activation function*

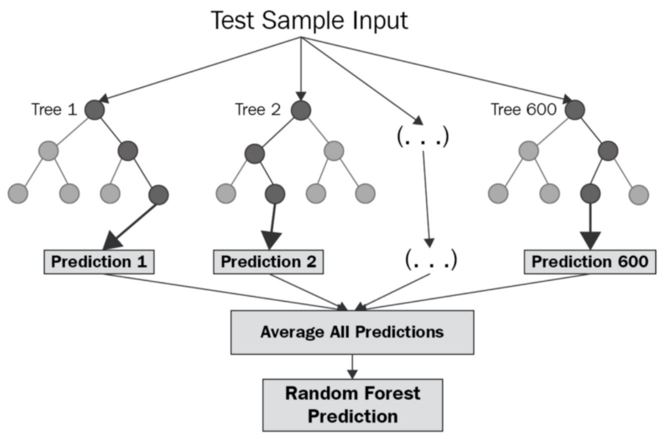
-tanh hyperbolic tangent activation

b)Random Forrest Regression

Random Forest is an ensemble learning technique that builds several decision trees. For regression problems, it works by training each tree on a random portion of the data and averaging the results from each tree.

Mathematical Formulation:
The algorithm creates terminal nodes (leaves) that generate predictions by recursively splitting the data based on characteristics for each decision tree in the forest. The average of all the trees' predictions, in the case of regression, is the final forecast for a new data point.

Example Tree of Random Forest



c)Support Vector Machine

SVM is a supervised machine learning technique that may be applied to regression and classification problems. It looks for the hyperplane that best divides classes or forecasts a continuous result. SVM is impacted by support vectors, or data points that are near the decision border, and works well in high-dimensional domains.

Mathematical Formulation:
In a binary classification problem, maximizing the margin between the two classes is the SVM goal. The following represents the categorization decision function:

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right)$$

where:

- x is the input,
- N is the number of support vectors,
- α_i are the Lagrange multipliers,
- y_i are the class labels,
- K is the kernel function, and
- b is the bias term.

d)Linear Regression

By fitting a linear equation, linear regression describes the connection between the independent variables (X) and the dependent variable (y). By minimizing the sum of squared discrepancies between the observed and anticipated values, it seeks to identify the line that best fits the data.

Mathematical Formulation:

For a basic linear regression equations is:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where:

- y is the dependent variable,
- x is the independent variable,
- β_0 is the intercept,
- β_1 is the slope,
- ε is the error term.

e)Extreme Gradient Boosting(XGBoost)

Extreme Gradient Boosting, or XGBoost for short, is an ensemble learning technique that builds a powerful predictive model by aggregating the outputs of several weak learners, typically decision trees. It fixes mistakes from earlier models and constructs trees one after the other.

Mathematical Formulation

In a regression problem, the objective function of XGBoost is the product of a regularization term and a loss function:

$$\text{Objective} = \sum_{i=1}^{n}\left[l(y_i, \hat{y}_i) + \sum_{k=1}^{K}\Omega(f_k)\right]$$

-n is the number of training instances

y_i is true label for the i-th instances

$y^{\wedge}i$ is the predicted label for i-th instances

l is the loss function measuring the difference between yi-y^i

K *is the number of trees*

f_k *is the k-th tree,*

Ω is the regularization sum.

V. EVULATION METRICS

The performance of five different models was evaluated to create the necessary forecast data for the project. Some metrics have been used for this purpose. The features of these metrics are as follows:

Mean Absolute Error (MAE): MAE calculates the average of the absolute differences between the model's predictions and the actual values. It measures the deviation of the model's predictions from the actual values. Lower MAE values are associated with better predictions, and it is suitable for assessing the extent of deviation between the model predictions and actual values in this project.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i |$$

Mean Squared Error (MSE): MSE computes the average of the squared differences between the predicted and actual values. It emphasizes larger error values and might be sensitive to outliers. This metric can be used to determine the extent of potential errors in the model's predictions.

$$\text{MSE} = \overset{\text{Mean}}{\boxed{\frac{1}{n}\sum_{i=1}^{n}}} (\overset{\text{Error}}{\boxed{Y_i - \hat{Y}_i}})^{\overset{\text{Squared}}{\boxed{2}}}$$

Root Mean Squared Logarithmic Error (RMSLE): RMSLE measures the logarithmic transformation of the mean squared differences between the logarithm of actual values and predictions. It is commonly used when dealing with a wide range of values in the target variable. RMSLE provides a more interpretable measure, particularly in scenarios where variables span several orders of magnitude.

$$RMSLE(y, \hat{y}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(y_i + 1) - \log(\hat{y}_i + 1))^2}$$

R-squared (R^2 - Coefficient of Determination): R-squared represents the proportion of the variance in the dependent variable that can be explained by the independent variables. Higher R-squared values indicate a better fit of the model to the data. This metric can help in understanding how well the model explains the variability in the data.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

For the project to succeed, selecting the right metrics to evaluate the predictive ability is crucial. Metrics like MAE (Mean Absolute Error), MSE (Mean Squared Error), and RMSLE (Root Mean Squared Logarithmic Error) quantify the error between predictions and actual values. The R^2 metric has a different nature, aiming for a high value as its objective. Lower values of these (not R^2) metrics enable more accurate predictions, facilitating efficient identification of peak points such as energy consumption and enabling more effective system design. By prioritizing these metrics, the aim is to compare the predictive abilities of different models and select the best-performing model.

Using certain metrics, the predictive capabilities of the models were measured, and the metric results for the first dataset are provided in Figure 1.

Figure 1.Result of Dataset 2

	LSTM Dataset:1	Random Regressor Dataset:1	SVM Dataset:1
Traning MAE	0.255349	0.075932	0.143073
Training RMSLE	0.178668	0.054803	0.109821
Training R^2	0.581665	0.968934	0.775772
Training MSE	0.318016	0.023616	0.170457
Validation MAE	0.260827	0.188398	0.498815
Validation RMSLE	0.202471	0.150732	0.317232
Validation R^2	0.352037	0.792158	0.055289
Validation MSE	0.431021	0.138256	0.628416
Test MAE	0.188530	0.411855	0.141993
Test RMSLE	0.147815	0.255591	0.113013
Test R^2	0.521445	-0.746864	0.683381
Test MSE	0.110112	0.401939	0.072851

The models' performances significantly differ based on the metric results in the table. Some observations drawn from these results are as follows:

The Random Regressor model showed better results compared to other models on the training and validation datasets. However, the negative R^2 value of the Random Regressor model on the test data indicates that the model lacks the ability to predict data as desired and generates random results.

The Support Vector Machine model demonstrated high performance on the training and test datasets. However, the results on the validation dataset are inadequate, indicating the weak overall predictive ability of the Support Vector Machine model.

The LSTM model's metric results are more reasonable compared to the other models. The LSTM model exhibits balanced performance across all three datasets, producing acceptable predictions.

Following the interpretation of these metrics, it is evident that the best choice among the models is the LSTM model. The LSTM model can generate more stable results compared to other models and demonstrate a more robust performance in handling overfitting or outliers in the dataset.

In the second dataset, three different models have been utilized. The metrics used to interpret the predictive capability of this model are presented in Figure 2.

Figure 2.Result of Data Set 2

Model	LSTM	XGBoost	Linear
Training			
MAE	0.1890	0.01536	9.812e-17
MSE	0.0561	0.00064	1.6445e-32
Validation			
MAE	0.2624	0.09641	0.25178
MSE	0.0929	0.01402	0.06911
Test			
MAE	0.2917	0.00580	0.44421
MSE	0.1224	0.04293	0.21169

Based on the metrics in the table, the following observations are made about the models:

MAE and MSE metrics serve as measures of how erroneous the model's behavior might be, with lower values being preferable. The Linear Regression model demonstrates near-perfect results, close to zero, on the training data, indicating highly effective learning. However, higher ratios in the validation and test metrics imply an overfitting issue in the model.

The LSTM model exhibits balanced error rates, albeit higher in comparison to the XGBoost Model. Consequently, it is inferred that XGBoost performs better for the second dataset. This is believed to be due to XGBoost's more efficient handling of irregular data, resulting in improved predictive capabilities.

In summary, while Linear Regression achieves near-perfect learning on the training data, it faces overfitting issues in the validation and test sets. On the other hand, although the LSTM model maintains balanced error rates, XGBoost outperforms it in handling more irregular data in the second dataset. XGBoost has been selected as the best model for this dataset.

The literature study referenced as [1] achieved an R^2 ratio exceeding 95%. However, it should be noted that this study trained the system using weather data. There's a possibility that correlation values among weather data might be higher than those among household appliances.

VI. RESULTS AND DISCUSSION

After selecting the LSTM model, an optimization process has been initiated to accurately predict the peak times of energy consumption. This necessitates the alteration of how flexibly classified home appliances operate at predicted peak points. To achieve this goal, it is crucial to assess how successful the flexible appliances are at capturing peak points and the extent to which overall energy consumption will be affected by the intended changes. The graph representing this objective is depicted in Figure 1.

Figure 1. Catching Point for Flexible Appliances

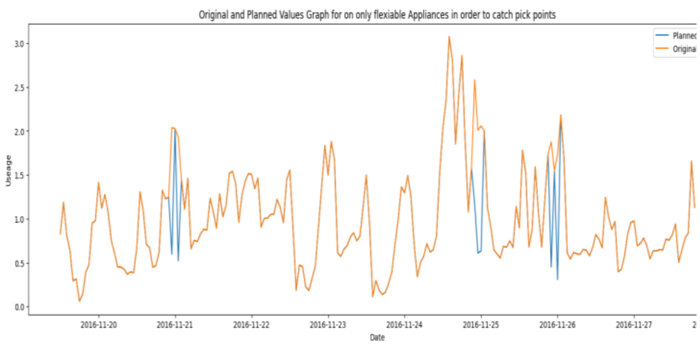
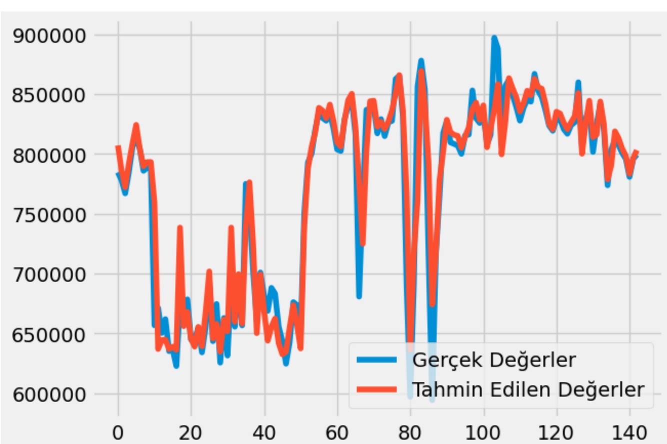


Figure 1 depicts the successful identification by the LSTM model of certain peak moments, showcasing the adjustment of flexible appliances to avoid these specific time frames. While the model operates partially as intended, it falls short in capturing certain high points accurately. This circumstance signals that the model's predictive capacity may not be exceptionally robust.

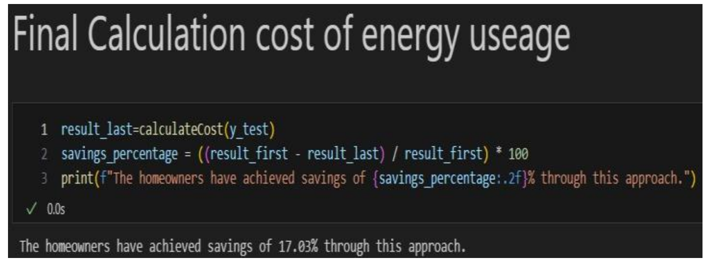
One of the most crucial criteria in determining the success of the project is the measurability of predictive capability. Figure 2 has been used to represent the predictive ability of the XGBoost model against the test data.

Figure 2. XGBoost Model Prediction Graph



As evident from the graph, XGBoost has acquired a significantly strong predictive ability, showcasing consistent success across the training, validation, and test datasets.

Figure 3.Cost of Energy Usage



After the selection of the LSTM model and optimization, alterations made in the initial dataset resulted in some changes in energy consumption and the bill, contingent upon the pricing function. The primary aim of these modifications is to yield a noticeable saving on the bill. Figure 3 depicts the output of the function that represents the gain ratio achieved in the project. According to this output, approximately a 17% saving has been attained on the homeowner's bill.

Systems actively used in the literature achieve a gain rate of around 28% [2]. In comparison,the project lags behind the existing systems in the literature in terms of gain percentage.

Thoughts and Discussions About Hybrid Systems in Smart Home

In the literature review focuses on three different systems, outlining their advantages and disadvantages. It's acknowledged that building ANFIS(Adaptive neuro-fuzzy systems) and ANN fuzzy systems (Neuro-fuzzy controllers) is a complex process.The construction process for FLC (Fuzzy Logic Controlers) systems is comparatively simpler than the other two. Nonetheless, identifying fuzzy clusters and rules that genuinely adapt for FLC systems demands significant experience. Due to the challenges posed by these systems, we aimed to evaluate them differently to design a new system for smart home systems. The main structure of system consists of three parts:

Initially, future predictions are derived utilizing an artificial intelligence model. Experimentation involved five distinct models: LSTM, XGBoost, Random Regressor, SVM, and Linear Regression. This prediction methodology presents a more streamlined predictive capability compared to ANN controllers (Neuro-fuzzy controllers) and ANFIS (Adaptive neuro-fuzzy systems) methods.

Subsequently, to navigate the challenges posed by fuzzy rules and clusters, three primary classifications have been delineated: flexible, semi-flexible, and non-flexible home appliances. The operational scheduling for flexible appliances entails pre or post-peak period arrangements, should such a period exist.

Semi-flexible appliance statuses may transition from standard to economic modes during a peak period. Non-flexible appliances lack such adaptability. Moreover, two crisis scenarios have been meticulously designed to ensure system functionality exclusively during peak consumption intervals. Any point identified below the top 3% of consumption or exceeding the peak point denotes the occurrence of a crisis scenario. Crisis scenario 1 encompasses optimization procedures for flexible and semi-flexible appliances. On the other hand, crisis scenario 2 pertains to consumption levels between the highest 3% and 6%, instigating optimization changes solely for semi-flexible appliances.

Finally, the optimization process integrates predictions from the most efficacious predictive model and incorporates the classifications of flexible, semi-flexible, and non-flexible appliances. This process also accommodates adjustments based on crisis scenarios to initiate pertinent optimization procedures.

A less complex system generally requires less time and resources. Designing, developing, and maintaining complex systems can take longer, whereas setting up and managing simpler systems can be quicker and easier. This can shorten the project completion time and reduce costs.

Understanding and managing simpler systems is often easier. The multitude of parameters, algorithms, or processes in complex systems can make detecting errors difficult and complicate the troubleshooting process. A simpler system, working with fewer variables, may make it easier to detect and rectify errors.

Moreover, simpler systems tend to be more flexible and scalable. While complex systems may consist of layered or interconnected components, simpler systems can be more modular and easily expandable or adaptable to changing needs.

Avoiding the design of long and complex smart home systems in favor of designing a simpler-level system can reduce the need for specialized expertise in developing energy optimization projects. It can minimize repetitive tasks in programming and system design, accelerating the development process. This, in turn, allows for the quicker completion of more extensive or intricate projects.

While recognizing the positive attributes of this system, it's crucial to assess its limitations. One such challenge is its adaptability to varying scenarios. The system faces difficulties in automatically adjusting to changes, especially when new appliances are introduced or removed from the household. In comparison, ANFIS demonstrates a propensity for yielding more effective outcomes in optimization tasks due to its adaptable learning of data mining processes.Therefore, systems developed with ANFIS exhibit a high level of success and demonstrate better adaptability to changing conditions.

The Main Contributions

We achieved an improvement of approximately 17 percent in addressing homeowners' high electricity bill problem. For this specific issue, we evaluated the performance of 5 different models with 2 different datasets and 4 different error metrics. We defined categories named flexible, semi-flexible, and non-flexible, and explained how these can be utilized for optimization problems.

VII. Conclusion

In this project, we examined the issue of energy efficiency optimization for smart homes. We addressed the problem of increasing energy consumption with the rise of smart home devices with two different solutions: producing more energy or consuming less energy. While working on this project, we reviewed old and new applications that work on this problem. At the end of this study, we decided that it is more appropriate to use data mining techniques such as Random Forest Regressor, Linear Regression, XGBoost, Support Vector Machine and LTSM. We applied 3 different models for each dataset we searched on the internet and observed the results. Thanks to the experiments, we concluded that the program works effectively. If we compare the results and models we obtained with current applications, we can say that theown model is much easier to design and implement, but it may lag behind some research on energy optimization. Based on these results, the study can be used as a recommendation system instead of applications generally used as autonomous systems. The project reduces energy use by capturing the peak points of flexible devices, but the reason why we got lower results compared to autonomous systems was that the models were not fully successful in capturing all peak points. Among the models, XGBoost and LSTM models gave the best results. The most important reason why we got good results in the XGBoost model was the application of advanced pre-processing processes and feature creation processes. The LSTM model was advantageous compared to other models because there was no overfitting. Overall, approach has shown promising results in improving energy efficiency in smart homes and can be used as an information or recommendation system through communication channels.
In conclusion, the research contributes to ongoing efforts to optimize energy consumption in smart homes. By leveraging data mining techniques and considering the flexibility of devices, we can make significant progress in reducing energy waste and creating more sustainable living environments.

VIII. Future directions

In this regard, it seems that deep learning applications can create more efficient results because generally used data mining techniques are very easily overfitting and dealing with this can lead to serious problems. Therefore, in future studies, using neural network structures and changing their capacity, as well as using regularization techniques against the overfitting problem, will yield much better results. For this project, an agreement can be made by contacting the regions that have switched to smart home systems and we can try to obtain more data sets because this is a new field with not many working examples. Apart from the dataset collected in this field, only datasets related to the electricity consumption of cities or countries can be found, and similar studies can be carried out with these datasets as a reference for future studies. As used in current applications, the rules set depending on the priority order and model, such as fuzzy logic, can be expanded and a more mature application suitable for human use can be created. If such steps are taken and progressed, we, as humans, will be able to adapt very easily to the smart home systems and energy optimization that will come into our lives.

References

[1] Candanedo, L.M.; Feldheim, V.; Deramaix, D. Data driven prediction models of energy use of appliances in a low-energy house. *Energy Build.* **2017**, *140*, 81–97.

[2] Mehmood, F.; Ahmad, S.; Ullah, I.; Jamil, F.; Kim, D. Towards a dynamic virtual iot network based on user requirements. *Comput. Mater. Contin.* **2021**, *69*, 2231–2244. I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] Shewale, A.; Mokhade, A.; Funde, N.; Bokde, N.D. An overview of demand response in smart grid and optimization techniques for efficient residential appliance scheduling problem. *Energies* **2020**, *13*, 4266. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] Beraldi, P.; Violi, A.; Carrozzino, G. The optimal management of the prosumer's resources via stochastic programming. *Energy Rep.* **2020**, *6*, 274–280. M. Young, The Technical

[5] Cao, Z.; O'Rourke, F.; Lyons, W.; Han, X. Home Energy Management System Incorporating Heat Pump Using Real Measured Data. *Sensors* **2019**, *19*, 2937.

[6] Candanedo, L.M.; Feldheim, V.; Deramaix, D. Data driven prediction models of energy use of appliances in a low-energy house. *Energy Build.* **2017**, *140*, 81–97.

[7] Mehmood, F.; Ahmad, S.; Ullah, I.; Jamil, F.; Kim, D. Towards a dynamic virtual iot network based on user requirements. *Comput. Mater. Contin.* **2021**, *69*, 2231–2244. I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[8] Shewale, A.; Mokhade, A.; Funde, N.; Bokde, N.D. An overview of demand response in smart grid and optimization techniques for efficient residential appliance scheduling problem. *Energies* **2020**, *13*, 4266. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[9] Beraldi, P.; Violi, A.; Carrozzino, G. The optimal management of the prosumer's resources via stochastic programming. *Energy Rep.* **2020**, *6*, 274–280. M. Young, The Technical

[10] Cao, Z.; O'Rourke, F.; Lyons, W.; Han, X. Home Energy Management System Incorporating Heat Pump Using Real Measured Data. *Sensors* **2019**, *19*, 2937.

[11] Ramos, S., Soares, J., Cembranel, S., Tavares, I., Foroozandeh, Z., Vale, Z., & Fernandes, R. (2021). Data mining techniques for electricity customer characterization. *Procedia Computer Science*, *186*, 475–488.

[12] Lissa, P., Deane, C., Schukat, M., Seri, F., Keane, M., & Barrett, E. (2021). Deep Reinforcement Learning for Home Energy Management System Control. *Energy and AI*, *3*, 100043.

[13] Abdelfattah, Abassi, et al. "Energy consumption patterns and inter appliance associations using data mining techniques." *E3S Web of Conferences*, vol. 336, 2022.

[14] Kumar, Neeraj, et al. "Optimizing energy consumption in smart homes using Machine Learning Techniques." *E3S Web of Conferences*, vol. 387, 2023, p. 02002.

[15] Xiang, Lei, et al. "Prediction model of household appliance energy consumption based on machine learning." *Journal of Physics: Conference Series*, vol. 1453, no. 1, 2020.

[16] Park, Sanguk. "Machine learning-based cost-effective smart home data analysis and forecasting for energy saving." *Buildings*, vol. 13, no. 9, 2023, p. 2397.

[17] Almohammed, Mustafa, et al. "Data Mining and analysis for predicting electrical energy consumption." 2022 International Conference on Artificial Intelligence of Things (ICAIoT), 2022.

[18] Rodrigues, F., Cardeira, C., & Calado, J. M. F. (2014). The daily and hourly energy consumption and load forecasting using artificial neural network method: A case study using a set of 93 households in Portugal. *Energy Procedia*, *62*, 220–229. doi:10.1016/j.egypro.2014.12.383

[19] Hou, X., Wang, J., Huang, T., Wang, T., & Wang, P. (2019). Smart Home Energy Management Optimization Method considering energy storage and Electric Vehicle. *IEEE Access*, *7*, 144010–144020. doi:10.1109/access.2019.2944878

IEEE