



## Article

# Predictive Analytics for Energy Efficiency: Leveraging Machine Learning to Optimize Household Energy Consumption

Piotr Powroźnik <sup>1</sup>  and Paweł Szcześniak <sup>2,\*</sup> 

<sup>1</sup> Institute of Metrology, Electronics and Computer Science, University of Zielona Góra, 65-516 Zielona Góra, Poland; p.powroznik@imei.uz.zgora.pl

<sup>2</sup> Faculty of Electrical and Computer Engineering, Rzeszów University of Technology, Wincentego Pola 2, 35-959 Rzeszów, Poland

\* Correspondence: p.szczeniak@iee.uz.zgora.pl; Tel.: +48-797-977-234

**Abstract:** This paper presents a novel machine learning framework useful for optimizing energy consumption in households. Home appliances have a great potential to optimize electricity consumption by mitigating peaks in the grid load or peaks in renewable energy generation. However, such functionality of home appliances requires their users to change their behavior regarding energy consumption. One of the criteria that could encourage electricity users to change their behavior is the cost of energy. The introduction of dynamic energy prices can significantly increase energy costs for unsuspecting consumers. In order to be able to make the right decisions about the process of electricity use in households, an algorithm based on machine learning is proposed. The presented proposal for optimizing electricity consumption takes into account dynamic changes in energy prices, energy production from renewable energy sources, and home appliances that can participate in the energy optimization process. The proposed model uses data from smart meters and dynamic price information to generate personalized recommendations tailored to individual households. The algorithm, based on machine learning and historical household behavior data, calculates a metric to determine whether to send a notification (*message*) to the user. This notification may suggest increasing or decreasing energy consumption at a specific time, or may inform the user about potential cost fluctuations in the upcoming hours. This will allow energy users to use energy more consciously or to set priorities in home energy management systems (HEMS). This is a different approach than in previous publications, where the main goal of optimizing energy consumption was to optimize the operation of the power system while taking into account the profits of energy suppliers. The proposed algorithms can be implemented either in HEMS or smart energy meters. In this work, simulations of the application of machine learning with different characteristics were carried out in the MATLAB program. An analysis of machine learning algorithms for different input data and amounts of data and the characteristic features of models is presented.

**Keywords:** machine learning; home energy management systems; smart appliances; household energy consumption profiles; elastic energy management



**Citation:** Powroźnik, P.; Szcześniak, P. Predictive Analytics for Energy Efficiency: Leveraging Machine Learning to Optimize Household Energy Consumption. *Energies* **2024**, *17*, 5866. <https://doi.org/10.3390/en17235866>

Academic Editors: Dimitrios Stimoniaris and Dimitrios A. Tsiamitros

Received: 31 October 2024

Revised: 18 November 2024

Accepted: 21 November 2024

Published: 22 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The electricity market is currently facing challenges resulting from significant legal and technical modifications implied by both regulators and market participants [1]. These challenges create a real opportunity for the effective implementation of mechanisms enabling the stimulation of electricity consumption management systems or shaping energy consumption through conscious behavior of recipients [2,3]. One of the tools enabling the implementation of this goal is dynamic electricity tariffs. This is a demand-side management technique that can reduce the peak load of electricity networks by setting different prices for electricity at different times depending on its supply and demand [4]. Peaks in the load profiles of the electricity system are the result of unregulated demand, resulting

mainly from the habits of electricity consumers, and a huge volume of generating capacity or other energy flexibility services is required to cover the peak load [5]. Often, this peak capacity of sources remains unused during off-peak periods, which results in a loss of profits and reduces the efficiency of the entire system. Additional services enabling the provision of capacity are also quite expensive and affect the prices of electricity. Dynamic pricing can influence electricity consumer behavior and shift energy consumption from peak to off-peak system load conditions.

In the classic structures of charging for electricity consumption in retail markets, flat prices (with fixed values) or block prices (the rate per unit of electricity either increases or decreases with the increase in the number of blocks of electricity consumption) were usually offered [6]. With flat fees, prices remained constant regardless of demand and energy production costs. This method of billing meant that the consumer of electricity was not interested in saving it, which resulted in the fact that the costs of generating electricity to meet peak demand were high compared to the costs outside the peak demand and did not reflect the actual costs of energy production and distribution. Block prices had a negligible effect on electricity consumption. The situation in the electricity market has changed significantly with the increase in the number of renewable energy source (RES) installations, energy storage facilities, and the emergence of prosumers. In addition, many countries are introducing laws requiring significant reductions in greenhouse gas emissions in various areas of life, including the transition to a low-emission energy system. Changes leading to a carbon-neutral energy system and the introduction of additional mechanisms for differentiating electricity prices will have a significant impact on consumer behavior. Consumer behavior should be supported by home energy management systems (HEMSs). The functionality of the HEMS should automatically plan the management of energy consumption by smart household appliances based on received signals about electricity production from RES, as well as signals about current dynamically changing energy prices. Changing the electricity consumption profile driven by energy prices for individual households is seen as a demand energy response service with great potential to reduce peaks in electricity demand. In addition to reducing peak demand, dynamic electricity tariffs also provide each consumer with the opportunity to reduce their electricity bill at a constant level of consumption by changing their energy consumption pattern (load shifting).

With the introduction of dynamic electricity tariffs, the operation of the HEMS will be determined mainly by price signals, and customers will react to oversupply or undersupply of energy in the electricity grid by changing their electricity consumption, thus supporting the stability of the grid. The costs of participation of households, including the loss of comfort of using electricity, changing habits, and financial costs, must be taken into account when building an HEMS [7–9]. Without the construction of a user-friendly HEMS system that is effective in reducing electricity bills, households may have significant barriers to accepting dynamic tariffs and participating in the creation of a friendly low-emission energy system. In the scientific literature, there are various review and research articles describing the impact of dynamic tariffs on consumer behavior or improving the stability of the power system. In [10], a systematic review of the literature on dynamic electricity pricing was conducted in order to understand the evolution of research in this area. Six main thematic areas of research on dynamic electricity pricing were presented: pricing scheme and modeling, price impact, user demand response, consumption scheduling and load scheduling technologies, cybersecurity threats, and fairness issues. In [11], the authors proposed three dynamic electricity tariffs. Then, using a demand response model of households with different annual electricity demands, different load-shifting capacities, and with or without the presence of a heat pump, the possible gross cost savings of households and CO<sub>2</sub> emission reductions were assessed. In [12], the benefits for different types of households resulting from increased automation, the choice of different electricity pricing options (dynamic tariffs), and the use of the possibility of participating in demand response programs were investigated. The results show that automated households with HEMSs

have lower electricity bills. The literature on household energy management, as presented in [13], highlights the significance of optimizing energy consumption in residential settings. For instance [13], delves into the intricate problem of optimizing the operation of a multi-energy building microgrid under various uncertainties.

In paper [14], the problem of dynamic pricing and scheduling of energy consumption in a microgrid is investigated, where the service provider acts as an intermediary between the utility company and the customers, buying electricity from the utility company and selling it to the customers. The service provider uses dynamic pricing to manage the microgrid. The authors propose a reinforcement learning algorithm that allows each service provider and customer to learn about their energy management strategy, including costs. The proposed energy scheduling algorithm enables them to reduce system costs due to the learning ability of both the provider and each customer. The proposed approach considers the specific problem of intermediating energy consumption by the service provider that also involves the customer, but does not take into account renewable energy sources also owned by the energy consumers. The article does not take into account HEMS control on the consumer side. An online reinforcement learning approach for specific recipients such as electric vehicle charging stations was proposed in the article [15]. The large share of electric vehicles as energy recipients is related to the rapid increase in demand for electricity causing load peaks in the power grid. Establishing an online pricing strategy is one way to solve the problem of increased energy consumption resulting from the use of EVs. The algorithm provides information on energy prices for a specific group of recipients, optimizing the operation of the network and the profits of a given electric vehicle charging station. This is not an algorithm for optimizing customer profits.

The use of the flexibility potential of the residential sector to modify electricity consumption patterns using appropriate pricing policies and HEMSs was the subject of, among others, the paper [16]. This paper presents a reinforcement learning (RL) approach to a price-based demand response (DR) program, where heating electrical appliances (EC) were the main controlled device. The results show the effectiveness of the proposed DR program, maximizing the aggregator's profits and satisfying the HEMS needs, while maintaining the system constraints. This algorithm primarily optimizes the energy seller's profits, taking into account the system constraints, and only indirectly affects the financial burden on electricity consumers. In the next article [17], the application of a smart meter with IoT technology and machine learning techniques is presented. The implemented machine learning algorithms are proposed to monitor and predict energy consumption. In this study, the authors estimate the electricity consumption of household appliances in an apartment using machine learning techniques, which are then used in demand management algorithms. Additionally, the proposed IoT system allows the consumer to remotely view the individual energy consumption indicator and possible prediction of energy consumption. The article does not link the management algorithm with dynamic prices. A case study of an HEMS for a two-story detached house in Japan with a dynamic pricing model is presented in the paper [18]. The next-day energy consumption forecast is developed based on historical data using a particle swarm optimization regression vector machine algorithm. Additionally, a dynamic pricing model is developed to guide the users' electricity consumption behavior and consider the power grid constraints. Three pricing schemes are proposed to optimize energy prices and consumer behavior. The proposed approach does not continuously determine the energy cost information but presents general pricing scenarios and their benefits. In [19], a deep-learning-based adaptive dynamic programming algorithm (ADPA) was presented to integrate real-time pricing with demand-side energy management optimization for microgrids. Machine learning methods are increasingly employed in short-term load forecasting for distribution transformer supply zones, particularly in the context of growing data privacy concerns. The research presented in [20] serves as an excellent example.

The presented brief review of previous research works indicates that the common goal of the works was to manage energy in such a way as to reduce the load on the

power grid and also to optimize its operation in the context of operating costs. The indicated publications also indicate the optimization of energy distributors' profits as the main goal, taking into account the limitations of the power grid. Information on energy costs on the consumer side is indirect information, which is intended to force changes in consumer behavior. The literature review shows a gap in knowledge, indicating the lack of publications clearly indicating electricity consumers as the main beneficiaries of the applied machine learning techniques, which are intended to optimize the incurred costs for energy in dynamically changing energy prices. Additionally, the algorithms proposed in our article are intended to support the increase in self-consumption of energy from renewable energy sources and support the operation of the power system through additional information transmitted to HEMSs. The proposed machine learning techniques do not yet take into account the use of battery energy storage.

The primary objective of this paper is to introduce a novel energy management solution, leveraging machine learning techniques, to empower users to minimize electricity costs. By providing timely recommendations, the HEMS aims to help users reduce energy consumption during peak hours and increase it during off-peak periods, especially when RESs are available. This approach benefits both users and distribution system operators (DSOs) by optimizing energy consumption patterns and reducing peak demand. Our research addresses a critical knowledge gap by developing an algorithm that prioritizes cost minimization for electricity consumers. This algorithm operates under dynamic energy pricing conditions, taking into account both historical user behavior and real-time renewable energy generation. While the primary focus is on reducing energy costs, this approach indirectly contributes to a more sustainable energy system by promoting increased self-consumption of renewable energy and reducing peak load demand on the power grid. By raising user awareness of energy consumption patterns and providing actionable insights, this solution can help foster a low-carbon future. The results presented in the paper indicated that the proposed approach achieved a dynamic balance between electricity supply and demand considering peak shaving and valley filling problems, and improved the rationality of the energy management strategy, thus ensuring the stable operation of the microgrid.

This paper is organized as follows: Section 2 provides a detailed description of the research problem, including a thorough examination of the research question, its theoretical foundations, and context. Section 3 justifies the choice of a specific machine learning model or algorithm and describes the data preparation process. This chapter also presents a detailed description of the model implementation, including the selection of libraries, tools, and programming environment. Section 4 describes the conducted simulations, including detailed descriptions of the simulation experiments, presentation of the results, and analysis of the results. The final chapter summarizes the most important findings and proposes directions for future research.

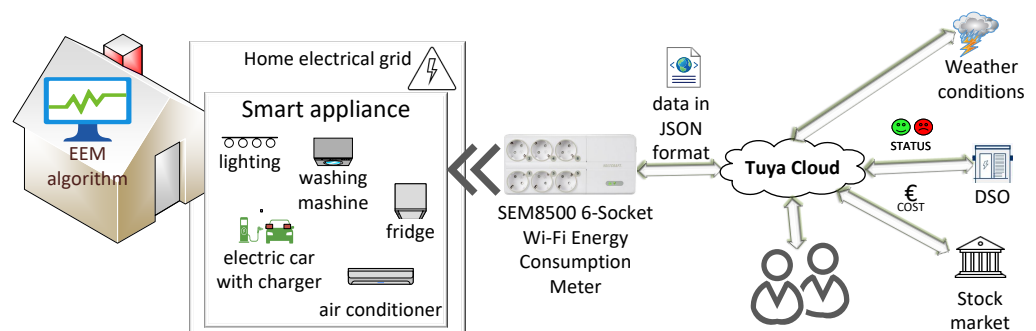
## 2. Description of the Research Problem

There are various ECs in an HEMS. These include conventional EC and smart appliances (SA). In the case of SAs, these appliances are characterized by the ability to adjust their power consumption level  $P_{SA}$  through control functions. This refers to the conscious actions of the user, e.g., by turning them on or off. It is also possible to change  $P_{SA}$  by changing the way the SA operates. This functionality can be used, for example, in an air heater, where there is more than one heating element. Air conditioners and heat pumps are other excellent examples [21]. The change in  $P_{SA}$  can also occur as a result of a change in the operating mode of the SA itself. Currently, eco modes are very often used in SAs, which reduce energy consumption. Additionally, SA manufacturers develop dedicated applications that allow for the management of individual ECs. In an HEMS, there are also a large number of ECs that do not have SA functionality. The lack of measurement and control functions means that the user himself is responsible for ensuring balanced energy

consumption. Such actions can be supported by the use of energy cost meters, which often include a remote control control function.

Despite the technological possibilities for the remote control of various ECs, users are reluctant to use such solutions. The problem is not the need to purchase additional ECs, but the fear of novelty and the need for education on new software and hardware solutions. For this group of users, the article will propose a solution that will aim to provide simplified messages. The messages will inform the user whether they should reduce energy consumption in the HEMS at a given moment (message = ‘reduce consumption’) or, on the contrary, increase energy consumption (message = ‘increase consumption’). It will also be proposed to leave the energy consumption at an unchanged level (message = ‘do not modify consumption’). Based on individual messages, the user will be able to consciously decide to start additional energy-intensive appliances or turn them off. Individual messages will be generated by the machine learning [22] algorithm presented in the article. In this case, the current price of electricity cost read from DSO, profiles of individual SA receivers, and conventional ECs connected to the electrical network using energy cost meters will be taken into account. The upper power threshold ( $P_{HEMS_H}$ ) and the lower power threshold ( $P_{HEMS_L}$ ) in the HEMS are also considered. The threshold value  $P_{HEMS_H}$  means the maximum power value that can be drawn by all ECs in a HEMS. In a HEMS, the power value drawn by all EC below the  $P_{HEMS_L}$  threshold may indicate a situation where energy from RES is available. In the absence of energy storage, this energy should first be used for the HEMS’s own needs. The proposed approach can be classified as a solution for elastic energy management (EEM) [23]. The user will be supported by machine learning to manage energy flexibly, taking into account the cost and power threshold for excessive or insufficient energy consumption. The wider application of such an approach in many HEMSs can also have a positive impact on DSOs due to the possibility of balancing the peak demand phenomenon [24] in power grids.

To verify the concept of applying machine learning to EEM, it was necessary to obtain data on real electricity consumption profiles in a sample HEMS. The structure of the data acquisition system is presented in Figure 1.



**Figure 1.** A system for collecting real-time energy consumption data from ECs within an HEMS.

Power consumption profiles of thirteen appliances were acquired using an Energy Consumption Meter SEM8500 6-Socket Wi-Fi [25] and transmitted to the Tuya Cloud [26] in JSON format [27]. The resulting data are presented in Figure 2.

The profiles presented in Figure 2 exhibit a wide range of energy consumption levels. The TV set consumed relatively low power over an extended period of approximately two hours. In contrast, the kettle drew a high power level only for the short duration required to boil water. The remaining appliances’ energy consumption patterns were associated with everyday activities such as cooking, laundry, and cleaning. Individual power consumption profiles  $P_{SA}$  were determined based on a dataset collected over a 30-day period. This allowed for the characterization of energy consumption profiles of a typical user in a single HEMS. Figure 1 also presents additional control signals that influence the cost. These include the electricity price for a given tariff, control signals from the DSO resulting from the need to counteract peak demand, and weather conditions. Weather conditions are



particularly significant due to their unpredictable nature and ability to directly impact the power grid infrastructure during natural disasters. For simulation purposes, example time intervals were assumed to reflect the cost of a given electricity tariff (Figure 3). The three adopted cost levels represent situations where the electricity price is: ‘cheap’, ‘normal’, or ‘expensive’.

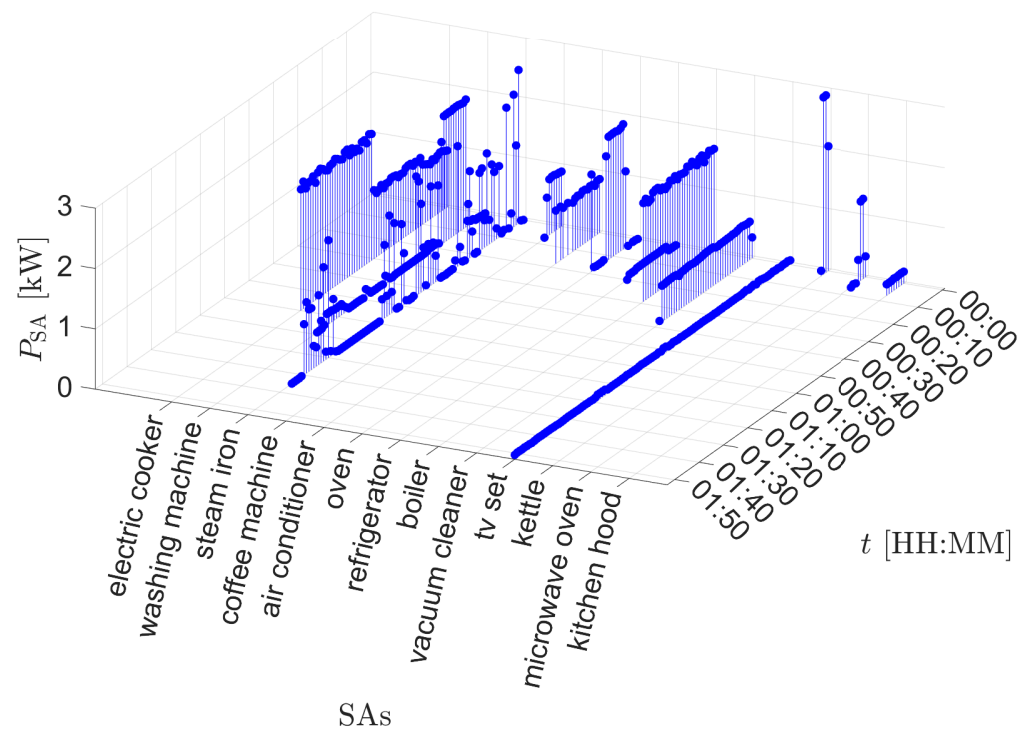


Figure 2. Power consumption profile  $P_{SA}$  of thirteen ECs.

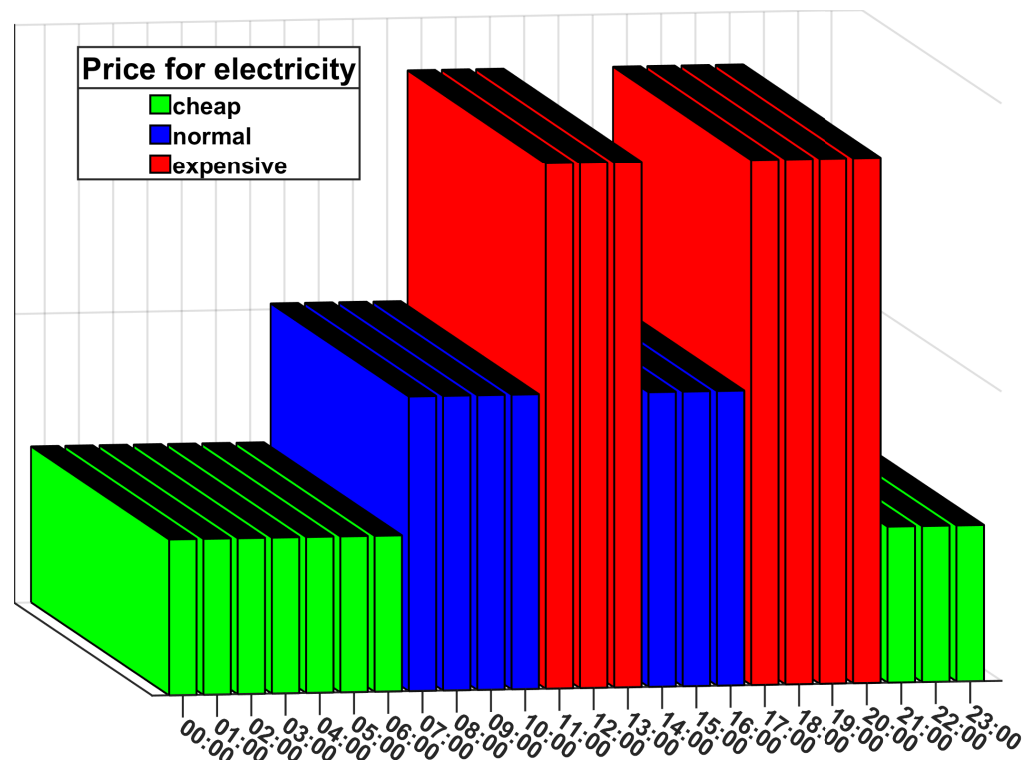


Figure 3. Temporal distribution of cost.

The cost distribution depicted in Figure 3 is designed to reflect the typical daily electricity consumption pattern. Peak demand typically occurs between noon and late evening, primarily due to increased user activity during these hours. To incentivize users to shift their consumption, the electricity price is increased during peak demand periods.

The proposed supervised machine learning algorithm (Algorithm 1) outlines the planned machine learning functionality for generating messages that will assist users in making decisions about their energy consumption behavior.

---

**Algorithm 1** Supervised machine learning algorithm for providing users with personalized recommendations on selecting one of the suggested energy consumption modes.

---

```

1: User Message Declaration (message): ‘increase consumption’, ‘reduce
   consumption’, ‘do not modify consumption’
2: Definition of electricity cost (cost): ‘cheap’, ‘standard’, ‘expensive’
3: Read off the lower power threshold in the HEMS ( $P_{HEMS_L}$ )
4: Read off the upper power threshold in the HEMS ( $P_{HEMS_H}$ )
5: repeat
6:   Read off the power consumed by the SAs ( $\sum P_{SA}$ )
7:   if  $\sum P_{SA} \geq P_{HEMS_H}$  then
8:     return message ▷ ‘reduce consumption’
9:   else if  $\sum P_{SA} \leq P_{HEMS_L}$  then
10:    return message ▷ ‘increase consumption’
11:   else
12:     if cost is ‘expensive’ then
13:       return message ▷ ‘reduce consumption’
14:     else if cost is ‘standard’ then
15:       return message ▷ ‘do not modify consumption’
16:     else if cost is ‘cheap’ then
17:       return message ▷ ‘increase consumption’
18:     end if
19:   end if
20: until  $\infty$ 

```

---

Algorithm 1 begins by initializing a set of user messages and defining thresholds  $P_{HEMS_L}$  and  $P_{HEMS_H}$ . Throughout its operation, the algorithm continuously monitors the total power consumption ( $\sum P_{SA}$ ) of all active ECs. This monitoring is essential to detect excessive power draw, which could potentially damage the electrical installation in the HEMS. If RESs are available, the algorithm will prompt the user to increase their energy consumption to avoid feeding excess power back into the grid. To reduce electricity bills, when the cost is ‘expensive’, the algorithm will encourage the user to decrease their power consumption. Conversely, when the cost is ‘cheap’, the user will be prompted to activate additional appliances. For time periods when the cost is ‘standard’, the user will not need to take any additional actions to adjust the power settings of their appliances.

### 3. Description of Supervised Machine Learning Implementation

In the following sections of this paper, we will present the details of the supervised machine learning implementation used to determine the user message. All operations involving training models to classify data were conducted under the assumption of a known set of input data (observations or examples) and known responses to the data (labels or classes). Simulation studies were carried out in MATLAB R2024b. A set of input data, along with its description, is presented in Table 1.

For input data conforming to the assumptions outlined in Table 1, cross-validation was performed with five folds to mitigate overfitting. Five-fold cross-validation [28] was chosen for its ability to balance model accuracy and computational cost. Dividing the data into five folds allows for a sufficient number of iterations to obtain a reliable model evaluation, while avoiding excessive computational time. Additionally, five-fold cross-

validation minimizes the risk of overfitting, as each subset of data is used for both training and testing. Supervised machine learning using various classifiers to categorize the input data from Table 1 was conducted for the machine learning models (*ml*) presented in Table 2.

**Table 1.** Input data and target variable specification for machine learning model.

Predictor Name	Description
time	Input data (observation) indicating the time at which the event occurred
kitchen hood, microwave oven, etc.	Input data (observation) indicating the power consumption values of an individual EC as shown in Figure 1
power sum	Input data (calculation) representing the total power consumption of all ECs at the time specified by time
energy price	Input data (calculation) allowing for the determination of the cost at the time specified by time based on the data presented in Figure 3
message	Response for input data considering Algorithm 1

**Table 2.** Machine learning models and their characteristics.

Machine Training Models	Description
<i>ml</i> <sub>1</sub>	Fine tree
<i>ml</i> <sub>2</sub>	Efficient logistic regression
<i>ml</i> <sub>3</sub>	Kernel naive Bayes
<i>ml</i> <sub>4</sub>	Linear support vector machine (SVM)
<i>ml</i> <sub>5</sub>	SVM kernel—a Gaussian kernel classifier for nonlinear classification of data
<i>ml</i> <sub>6</sub>	Boosted trees
<i>ml</i> <sub>7</sub>	Narrow neural network—a neural network classifier with one fully connected layer of size 10

The fine tree (*ml*<sub>1</sub>) model was selected due to its ability to handle both numerical and categorical data, its interpretability, and its capability to handle missing values. The efficient logistic regression (*ml*<sub>2</sub>) model was chosen for its speed and efficiency, especially with large datasets; its probabilistic outputs; and its interpretable coefficients. The kernel naive Bayes (*ml*<sub>3</sub>) model was selected for its ability to handle both continuous and categorical data, its relative simplicity and efficiency, and its capability to handle missing values. The linear support vector machine (SVM) (*ml*<sub>4</sub>) model was chosen for its effectiveness in high-dimensional spaces, its ability to handle complex decision boundaries, and its robustness to outliers. The SVM kernel—Gaussian kernel classifier (*ml*<sub>5</sub>) model was selected for its ability to handle non-linearly separable data, its effectiveness in high-dimensional spaces, and its robustness to outliers. The boosted trees (*ml*<sub>6</sub>) model was chosen for its high accuracy, robustness, ability to handle complex relationships between features and the target variable, and its capability to handle missing values. The narrow neural network (*ml*<sub>7</sub>) model was chosen for its ability to learn complex patterns, its high flexibility, and its capability to handle large datasets.

To preserve the original characteristics of the data, no data cleaning was performed, except for handling missing values using mean imputation. A new feature, power sum, was introduced to capture the total power consumption of all ECs (Table 1). All features were normalized using min-max scaling to improve model convergence. This approach allowed us to maintain the authenticity of the data while enhancing the performance of the models.

For model *ml*<sub>1</sub>, a classification tree was constructed using the MATLAB function `fitctree`. The Gini diversity index was employed as the splitting criterion, with the maximum number of splits set to 100. Surrogate splits were disabled. To build the *ml*<sub>2</sub> model, a decision tree classification algorithm implemented in MATLAB's `fitctree` function was employed. A `templateLinear` template was used to create a base model, defining a logistic regression classifier with automatic tuning of the regularization parameter `lambda`. Additionally, a beta change tolerance of 0.0001 was set. The final multi-class classifier,



classificationLinear, was trained on the training set using a one-versus-one strategy and predefined classes. A naive Bayes classification model ( $ml_3$ ) was constructed using MATLAB's `fitcnb` function. A normal kernel was selected, assuming a normal conditional distribution for features within each class. The Support parameter was set to Unbounded, allowing features to take on any value. Data standardization was applied to eliminate differences in scales between features. For the  $ml_4$  model in MATLAB, a linear SVM classifier was employed using the `fitcecoc` function. The SVM was configured with a linear kernel, automatic kernel scale, and a box constraint of 1. Data standardization was applied prior to training. A one-vs-one SVM classifier, implemented using MATLAB's `fitcecoc` function, was employed for the  $ml_5$  model. The SVM kernel was automatically selected, and the model was trained with a maximum of 1000 iterations. For the  $ml_6$  model, we employed MATLAB's `fitcensemble` function to train an AdaBoost ensemble classifier. The ensemble consisted of 30 decision trees, each with a maximum of 20 splits. The learning rate was set to 0.1. For the construction of the neural network model  $ml_7$  in MATLAB, the `fitcnet` function was employed with the following settings: the network had one hidden layer consisting of 10 neurons with a ReLU activation function; the training process was carried out using the stochastic gradient descent algorithm with an L2 regularization coefficient of 0; and the maximum number of iterations was set to 1000. Prior to training, the input data were standardized.

#### 4. Results of Simulation Studies

The initial machine training of models ( $ml$ ) to classify input data (Table 1) was conducted under the assumption of limiting the input dataset to a single day, instead of the full 30-day period for which HEMS data was collected (Figure 1). Moreover, in these experiments, no explicit data partitioning was performed to reserve a portion for testing. A comparison of the resulting classification accuracy for the message generated by the  $ml$  listed in Table 2 is presented in Figure 4.

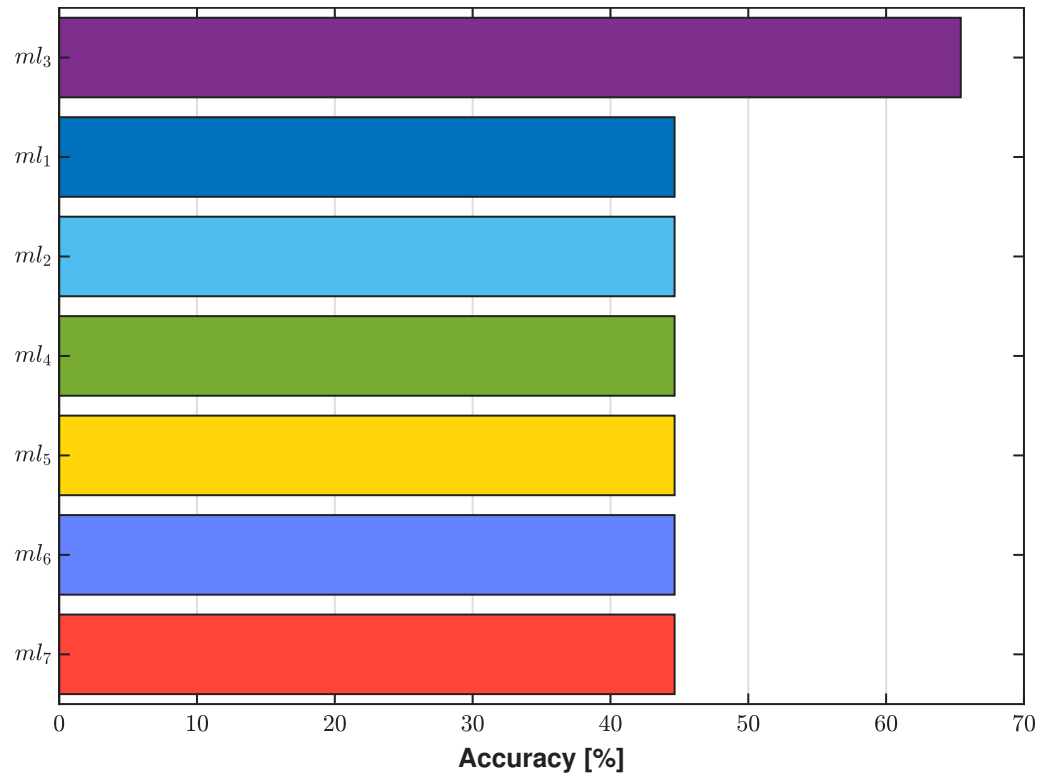
The highest classification accuracy for the message was achieved by  $ml_3$  (Table 2). The remaining six  $ml$  exhibited similar classification accuracy levels. To further analyze the best-performing model,  $ml_3$ , a validation confusion matrix was generated (Figure 5).

Based on the data presented in Figure 5, it can be observed that the currently selected classifier misclassified messages in each class. The most significant misclassification occurred when the true class was 'reduce consumption' but the predicted class was 'increase consumption'.

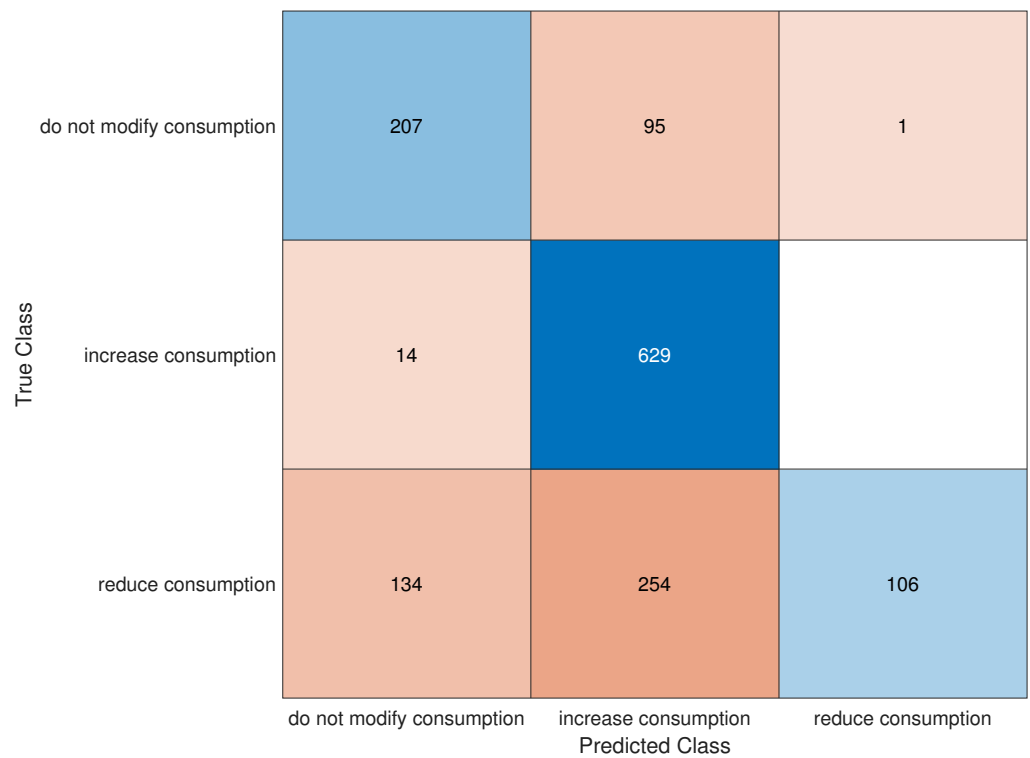
To investigate the impact of reserving a percentage of the data for testing on the classification accuracy of the message, the simulation experiments were repeated. A 10% portion of the input data was set aside for testing. The results of the validation confusion matrix for  $ml_3$  with 10% of the data reserved for testing are presented in Figure 6.

Allocating an additional 10% of the data for testing resulted in a degradation of the classification performance for message between the true class and the predicted class. This outcome is attributed to an insufficient amount of training data for  $ml$ . In subsequent simulation studies, the size of the training dataset was increased to encompass the full 30 days of data collected from the selected HEMS (Figure 1). The assumptions for these subsequent simulation studies are described in the test scenarios ( $sc$ ) outlined in Table 3.

For comparative purposes, Table 3 includes additional  $sc$  for a single day without ( $sc_1$ ) and with a test set ( $sc_2$ ). Scenarios  $sc_3$  to  $sc_6$  were designed to determine the minimum number of days required for each  $ml$  to achieve satisfactory classification accuracy for the message. The results of the accuracy experiments demonstrating the relationship between  $sc$  and  $ml$  are presented in Figure 7.



**Figure 4.** Comparison of message classification accuracy for one day: without a test set of data reserved for testing.



**Figure 5.** Validation confusion matrix for  $ml_3$  for a single day without a reserved test set.

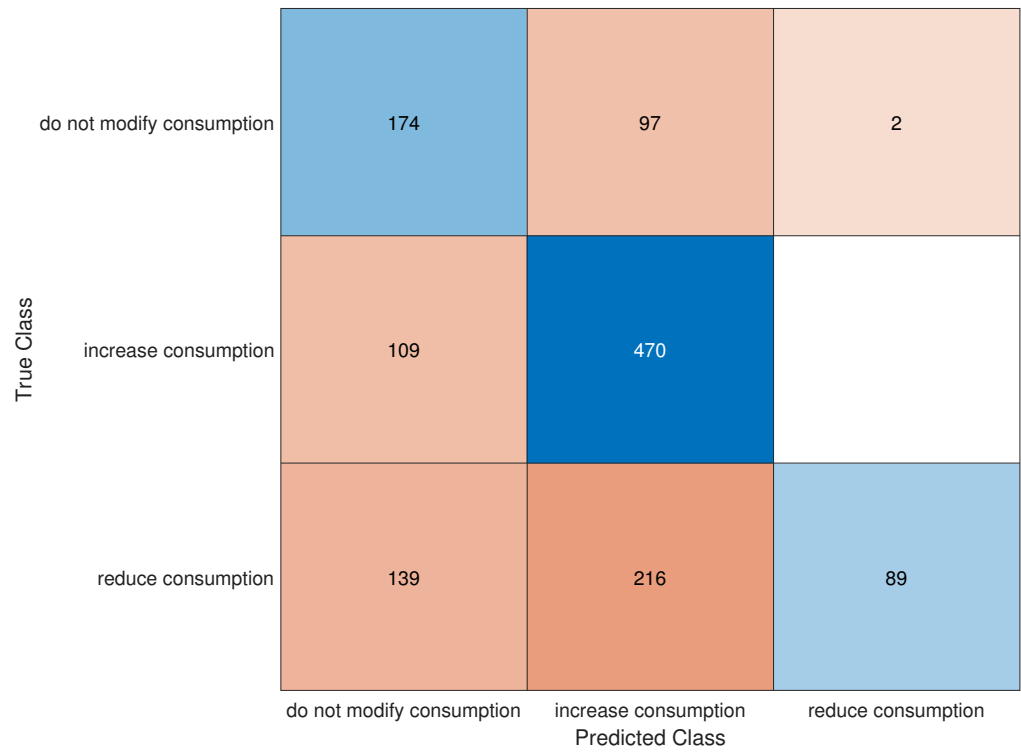


Figure 6. Validation confusion matrix for  $ml_3$  using a single day of data with a 10% test set.

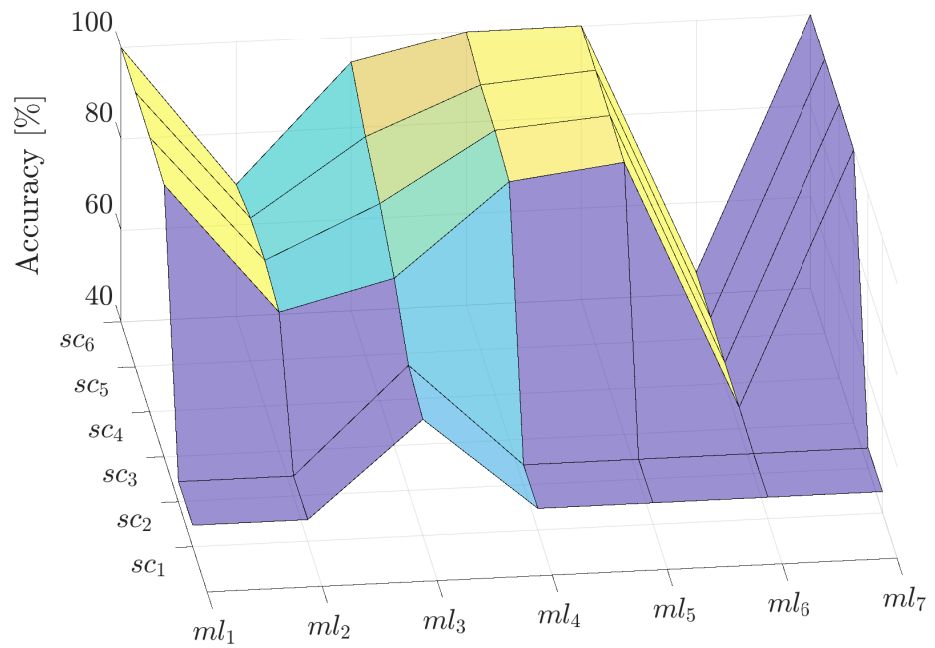


Figure 7. Simulation results for  $ml$  models, showing accuracy for different scenarios ( $sc$ ).

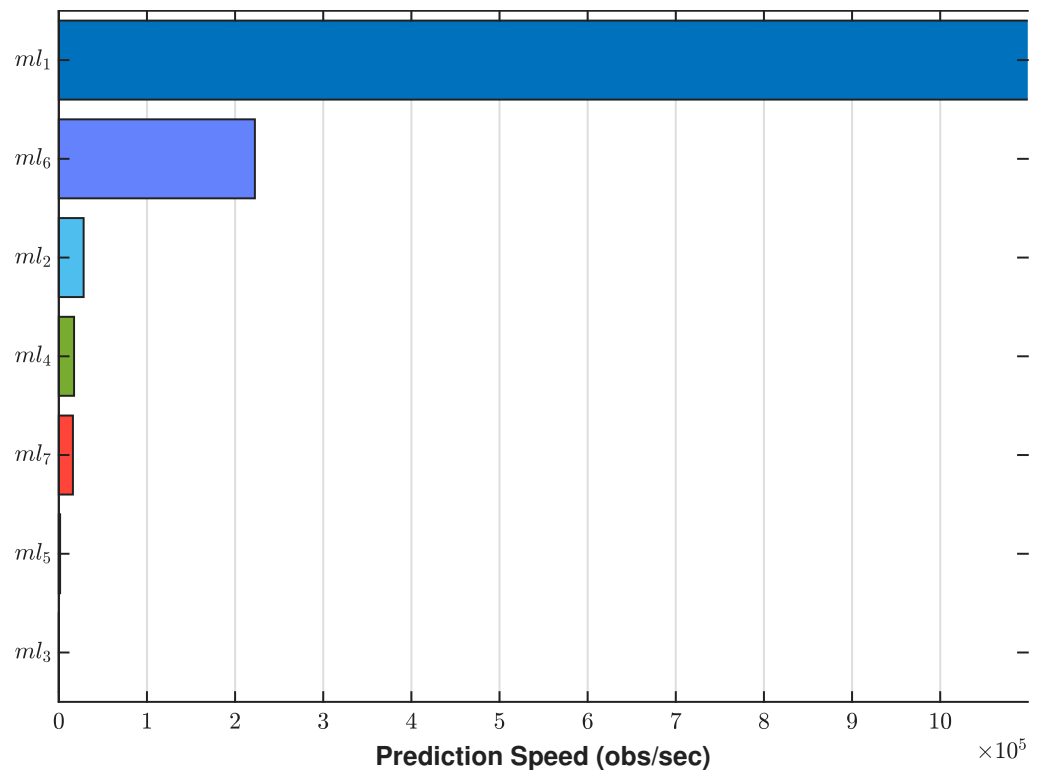
Table 3. Test scenario parameters for model evaluation.

$sc$	10% of the Known Set of Input Data for Training	Number of Days for Which the Known Set of Input Data Was Defined
$sc_1$	✗	1
$sc_2$	✓	1
$sc_3$	✓	3
$sc_4$	✓	7
$sc_5$	✓	14
$sc_6$	✓	30

Based on the accuracy results presented in Figure 7,  $ml_6$  exhibited the lowest classification accuracy for the message. Increasing the size of the training dataset in subsequent scenarios did not improve the accuracy for  $ml_6$ .  $ml_2$  showed slightly better performance, with an accuracy of around 70%. In contrast,  $ml_3$  achieved the highest accuracy for the smallest training dataset size ( $sc_1$  and  $sc_2$ ). However, increasing the training dataset size for  $ml_3$  led to an average improvement in accuracy of 5.8% in subsequent scenarios. For  $ml_1$ ,  $ml_4$ ,  $ml_5$ , and  $ml_7$ , an accuracy of approximately 100% was achieved as early as  $sc_3$ . In summary, for small training datasets,  $ml_3$  is the best choice for maximizing the classification accuracy of the message. For larger training datasets,  $ml_1$ ,  $ml_4$ ,  $ml_5$ , or  $ml_7$  are preferable.

To conduct a more comprehensive analysis of the individual  $ml$ , further investigations were carried out for test scenario  $sc_6$ . Test scenario  $sc_6$  was selected due to its having the largest input data size. For this purpose, the criteria of prediction speed (Figure 8), machine learning model size (Figure 9), and training time (Figure 10) were used for comparative analysis.

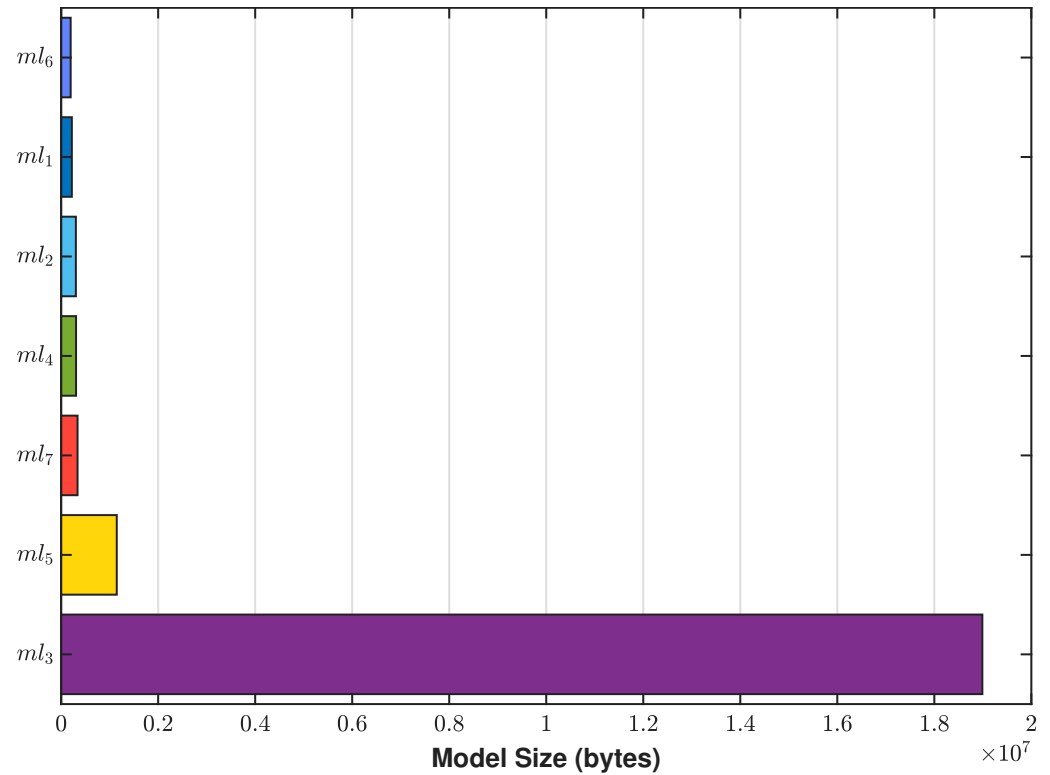
Prediction speed enabled the determination of the prediction speed in observations per second, a measure of the predictive model's efficiency. It indicates the number of observations a model can analyze and classify (or predict) in one second. Models  $ml_3$  and  $ml_5$  exhibited the lowest comparable prediction speed values (Figure 9). In contrast, model  $ml_1$  had a significantly higher prediction speed.



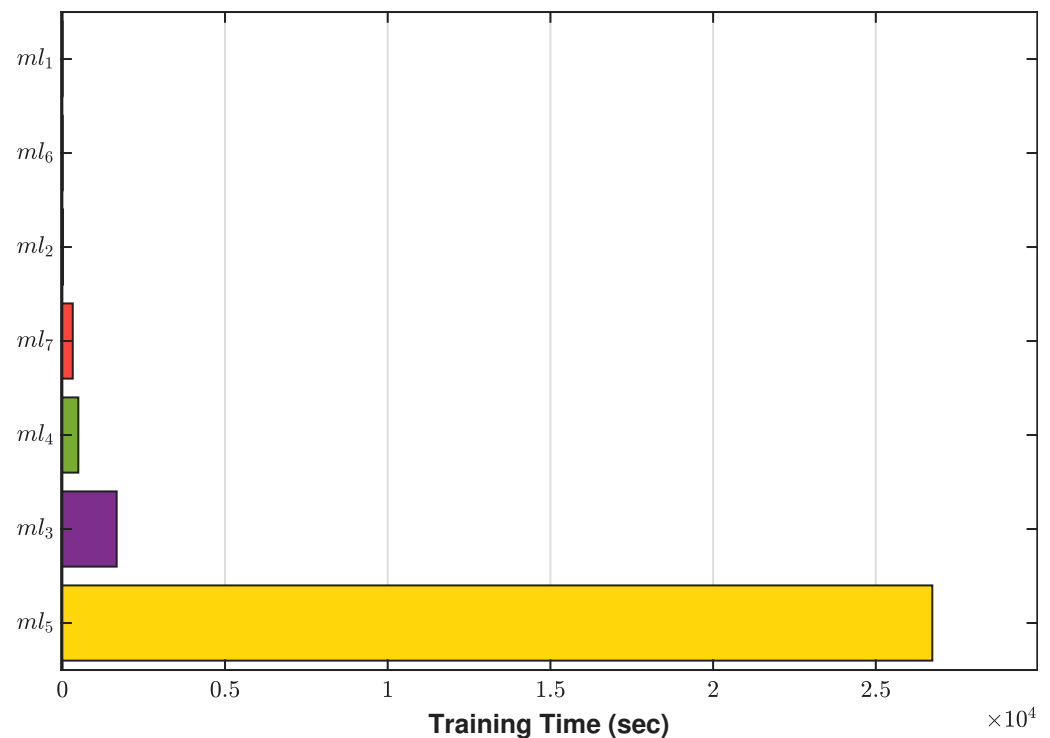
**Figure 8.** Dependence of prediction speed on model architecture in scenario  $sc_6$ .

Machine learning model size allowed us to determine the size of a given model, specifically the number of parameters the model must learn during training. Models  $ml_6$ ,  $ml_1$ ,  $ml_2$ ,  $ml_4$ , and  $ml_7$  had comparably smaller machine learning model sizes (Figure 9). Model  $ml_3$  had a significantly larger machine learning model size.

The training time parameter indicates the time it takes to train a machine learning model. It is a key performance indicator, especially for large and complex models. Models  $ml_1$ ,  $ml_6$ , and  $ml_2$  had the lowest training time (Figure 10). Model  $ml_5$  had a significantly higher training time.



**Figure 9.** Relationship between model size and computational performance.



**Figure 10.** Time required to train individual machine learning models.

To evaluate model performance, additional metrics were employed: precision, recall, and F1-score. Weighted averages were calculated for each metric, presented in Figures 11–13, respectively. The choice of weighted averages was motivated by the need to account for class imbalances in the dataset. This allowed for a more objective evaluation of the model, particularly in cases of significant class imbalance.



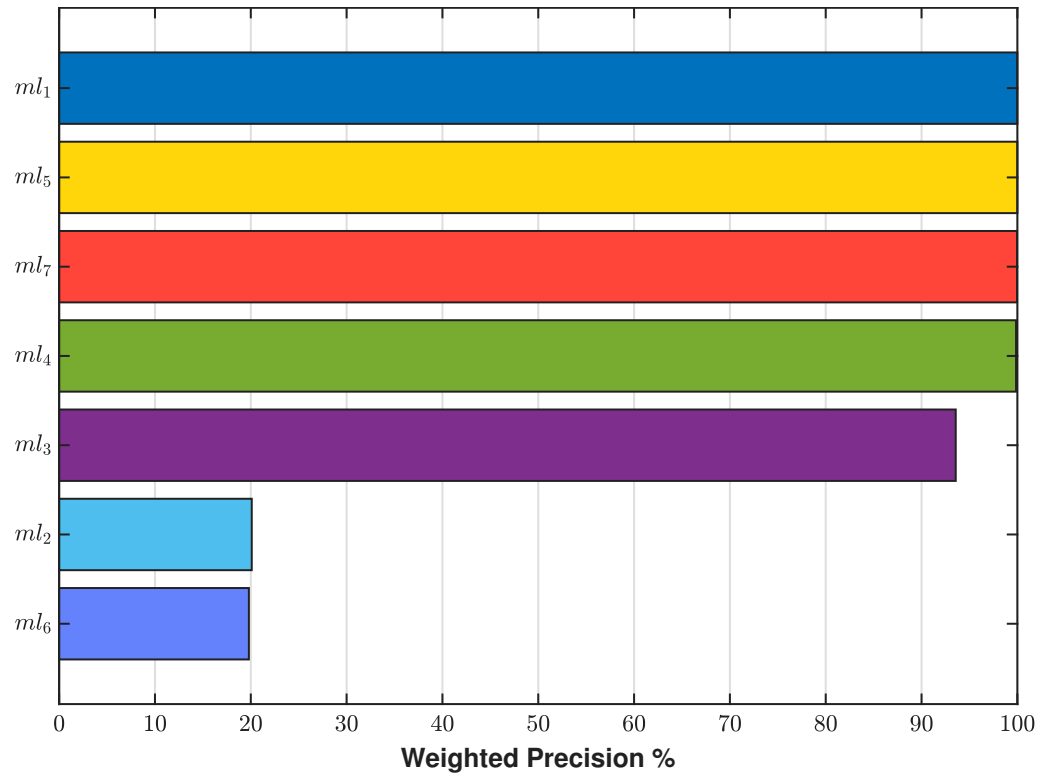


Figure 11. Weighted precision of the classifier for  $sc_6$ .

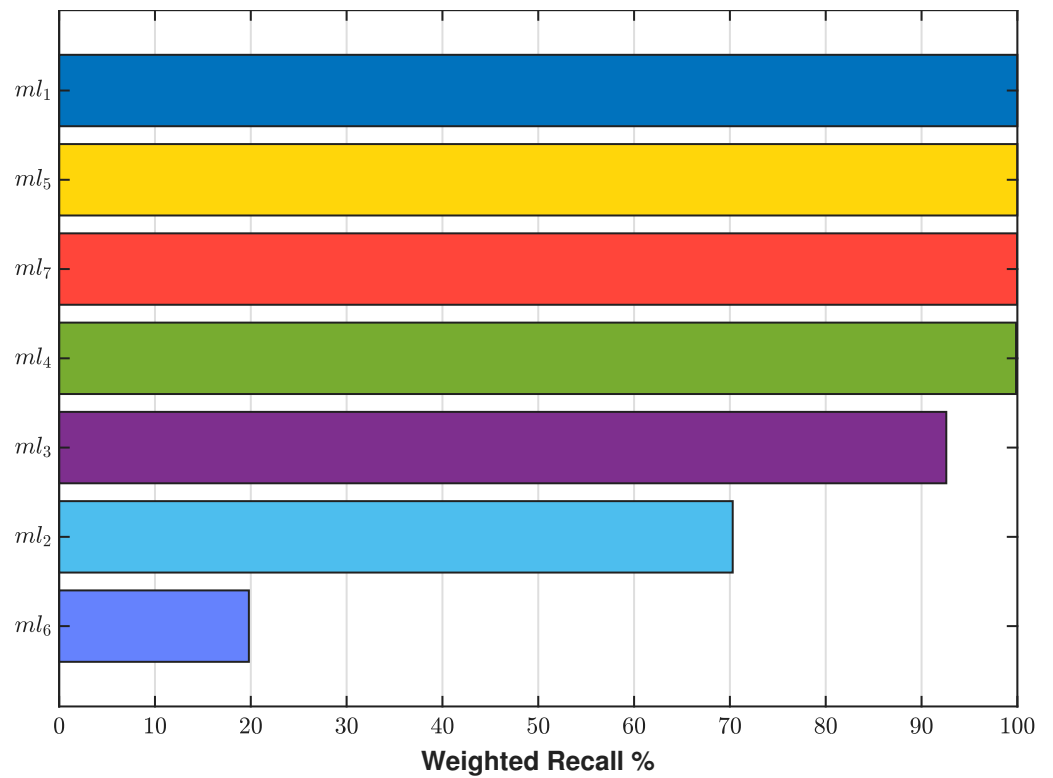
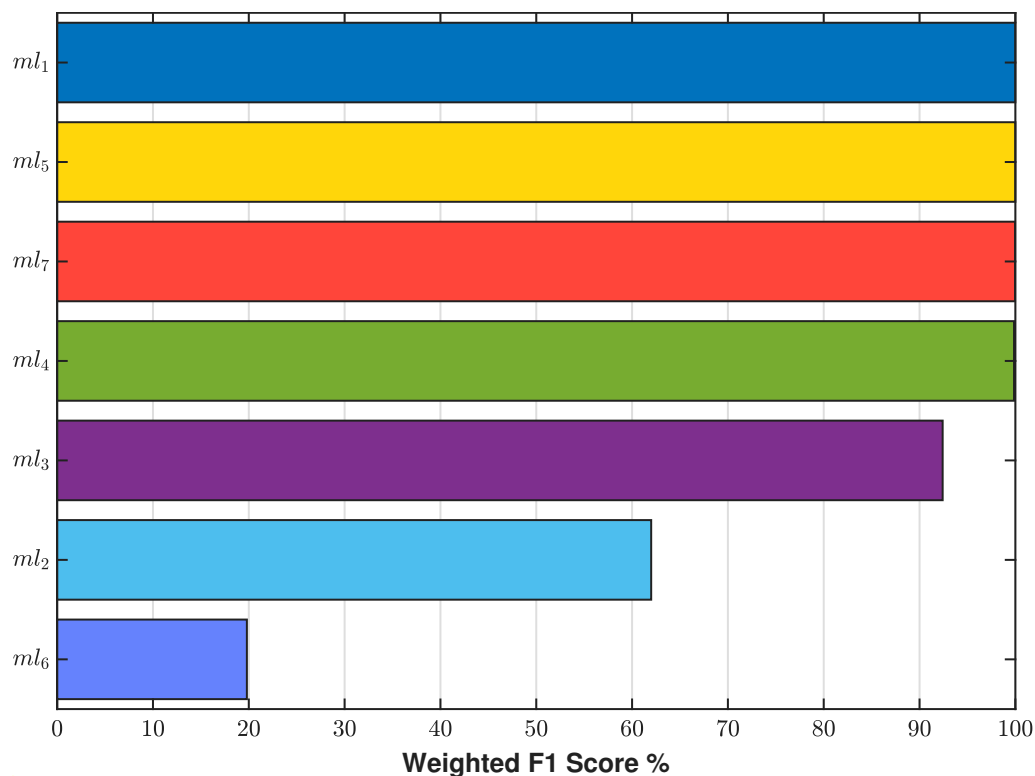


Figure 12. Weighted recall of the classifier for  $sc_6$ .



**Figure 13.** Weighted F1-score of the classifier for  $sc_6$ .

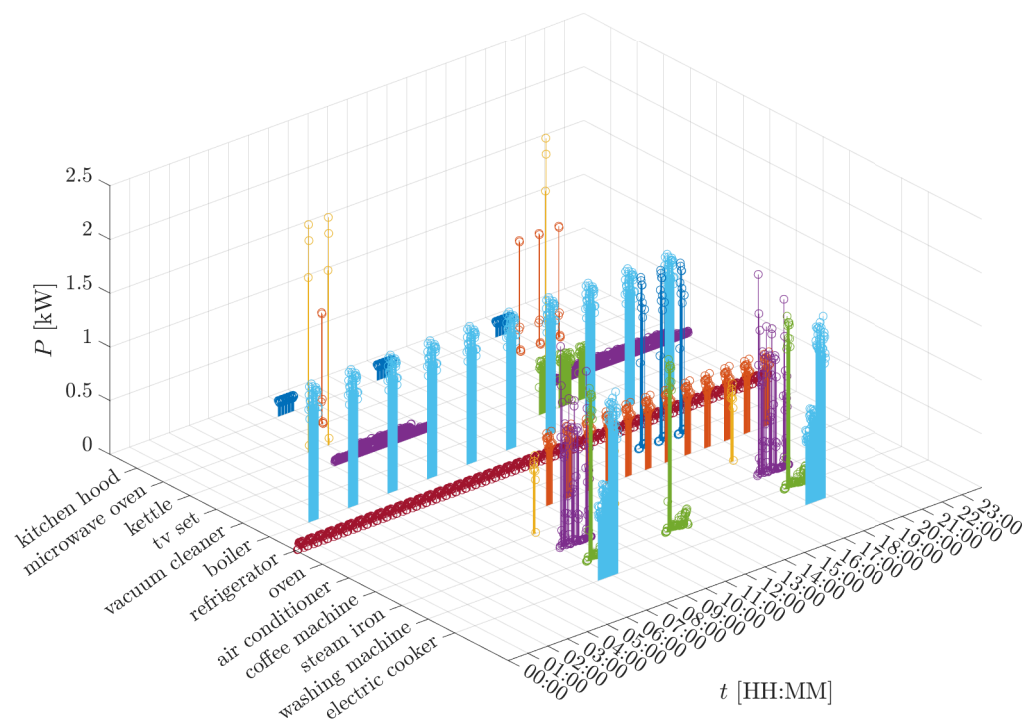
For models  $ml_1$ ,  $ml_5$ ,  $ml_7$ ,  $ml_4$ , and  $ml_3$ , a weighted precision value above 90% indicates a very high-quality model (Figure 11). This means that the model is highly effective in identifying true positive instances, even in datasets with imbalanced class distributions. In contrast, models  $ml_2$  and  $ml_6$  achieved a weighted precision of approximately 20%, meaning that only about 20% of the instances classified as positive were actually positive. The remaining cases were false positives, incorrectly classified as positive.

A weighted recall above 90% for machine learning models  $ml_1$ ,  $ml_5$ ,  $ml_7$ ,  $ml_4$ , and  $ml_3$  signifies that these models are highly effective in identifying all actual positive cases within the dataset (Figure 12). Models  $ml_1$ ,  $ml_5$ ,  $ml_7$ ,  $ml_4$ , and  $ml_3$  rarely miss positive instances. This implies that if a positive instance exists in the dataset, there is a very high probability that models  $ml_1$ ,  $ml_5$ ,  $ml_7$ ,  $ml_4$ , and  $ml_3$  will detect it. Since this is a weighted recall, models  $ml_1$ ,  $ml_5$ ,  $ml_7$ ,  $ml_4$ , and  $ml_3$  account for the weight of each class in the dataset. This means they are effective in identifying both frequent and rare positive classes. These models rarely misclassify positive instances as negative. A weighted recall value of 70% indicates that model  $ml_2$  is fairly good at detecting actual positive cases. Model  $ml_2$  less frequently overlooks instances that should be classified as positive. Model  $ml_6$ , on the other hand, misses many instances that should have been classified as positive. In other words, the model has a high number of false negatives. Model  $ml_6$  is unable to effectively detect the majority of cases belonging to the positive class.

Models  $ml_1$ ,  $ml_5$ ,  $ml_7$ ,  $ml_4$ , and  $ml_3$  (Figure 13) effectively identify true positive cases (high precision) while minimizing the number of false negatives (high recall). The F1-score, being the harmonic mean of precision and recall, indicates a good balance between these two metrics when its value is high. A high weighted F1-score suggests that models  $ml_1$ ,  $ml_5$ ,  $ml_7$ ,  $ml_4$ , and  $ml_3$  perform well on both frequent and rare classes, as it considers the weight of each class in the calculations. In contrast, a weighted F1-score of 60% indicates moderate performance for model  $ml_2$ . A low weighted F1-score means that model  $ml_6$  is unable to classify instances effectively. Model  $ml_6$  frequently misclassifies instances as positive (false positives) and misses many cases that should have been classified as positive (false negatives).

According to the evaluation metrics of precision, recall, and F1-score, models  $ml_1$ ,  $ml_5$ ,  $ml_7$ , and  $ml_4$  are the most suitable for test scenario  $sc_6$ . These models achieved high precision, recall, and F1-score, indicating a strong ability to accurately classify instances and balance between precision and recall.

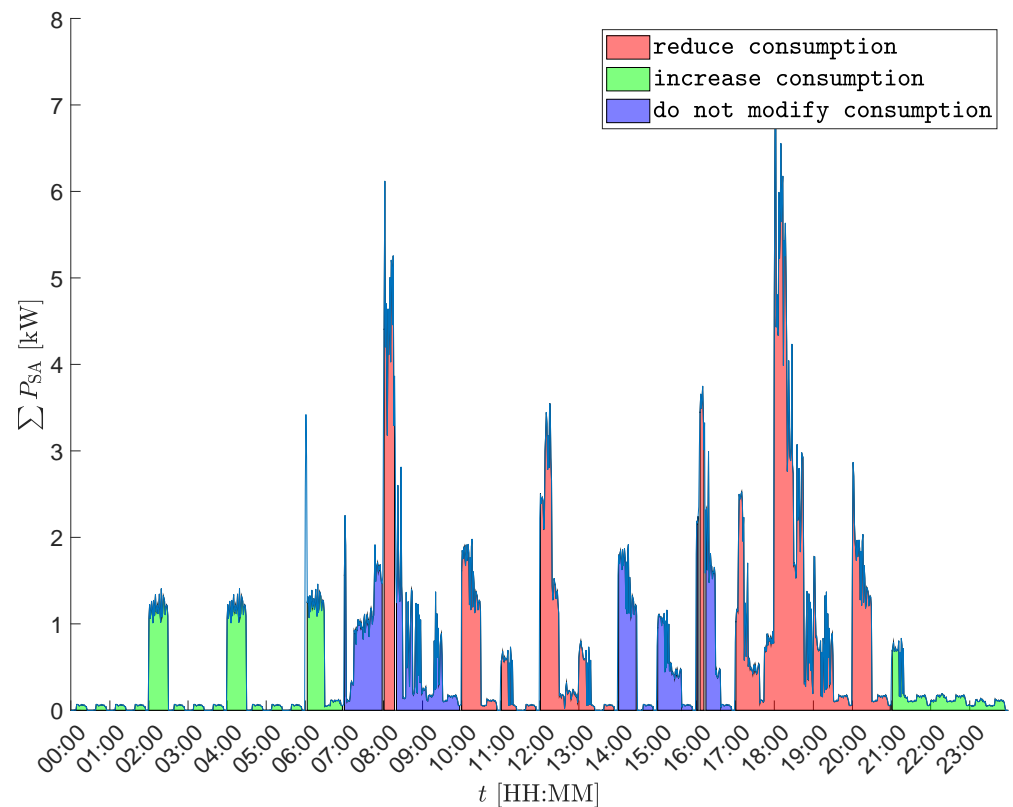
Based on the comparative analysis of previously presented data, machine learning model  $ml_1$  was selected due to its superior performance among the evaluated algorithms. A comprehensive case study was conducted on  $ml_1$ . The case study was conducted based on the results of implementing the  $ml_1$  within the MATLAB environment. The function `m1.predictFcn` was utilized to generate predictions. The case study involved a domestic setting where various energy-consuming SAs operated according to the schedule depicted in Figure 14. The schedule outlined the operation of energy-consuming SAs for household tasks such as cleaning and cooking. Additionally, a low, periodic energy consumption was assumed for appliances like refrigerators. Activities such as ironing and using the stove resulted in short-term spikes in energy demand. For the energy consumption profile illustrated in Figure 14, a case study was performed following the algorithm presented in Algorithm 1. Moreover, the analysis incorporated energy prices corresponding to the costs outlined in Figure 3.



**Figure 14.** Household energy consumption schedule for a case study.

Figure 15 illustrates the total power demand of all electrical SAs ( $\sum P_{SA}$ ) at any given moment in time ( $t$ ). The figure also highlights the specific time intervals during which the machine learning algorithm  $ml_1$  will deliver notifications messages to the user.

An analysis of the notifications messages generated (Figure 15) demonstrates a clear correlation with the energy costs presented in Figure 3. The machine learning algorithm  $ml_1$  leverages these cost data to provide the user with recommendations for appliance usage during off-peak hours, when electricity rates are lower. As outlined in the algorithm depicted in Algorithm 1, the content of these notifications messages varies at approximately 6:00, 8:00, and 16:00. During these time periods, the user is alerted to the possibility of exceeding the predefined power threshold ( $P_{HEMS_H}$ ) and is advised to reduce energy consumption accordingly.



**Figure 15.** Prediction of message indications for a user based on the  $ml_1$ .

## 5. Conclusions

This paper presents a novel approach to HEMS energy management based on supervised machine learning. Instead of automatically activating or deactivating appliances, the system provides suggestions in the form of notifications (message). Users are not required to analyze power consumption levels of individual appliances, current electricity prices, or the status of the power grid. Simulations were conducted to evaluate various  $mls$ . Input data were collected from electrical appliances in a sample household to create diverse  $sc$ . The  $mls$  were assessed based on the accuracy of their predictions and model complexity. The research also considered the impact of the machine learning model size, expressed in terms of the number of parameters, on its deployability on ECs with limited computational resources. For the selected  $ml$ , an analysis of of notifications (message) for specific events was conducted.

The main objectives of this work were: to optimize the multi-source power supply in the microgrid (increase the self-consumption of energy from local sources); to adapt flexible power sources in the microgrid (e.g. energy storage) in order to reduce the overall operating costs; and to achieve the goals of developing a sustainable energy network. This work does not consider the optimization of energy costs on the consumer side but only considers the optimization of the power system operation.

The proposed method of supporting the decision-making processes of electricity consumers in the article, taking into account the current load profiles and current generation of energy from renewable energy sources and information on dynamic prices, has a limited impact in the form of information on the possibilities of reducing electricity costs through its appropriate use. In the future, research is planned on including barrage energy storage in the decision-making process, as well as signals with information on short-term prediction of energy prices and prediction of generation from renewable energy sources. Additionally, research is planned to include the proposed machine learning algorithms in automated HEMSs. The last research proposal is behavioral studies of electricity consumers with algorithms taking into account the approach of notification of energy prices with automated

algorithms. This research should provide information on whether a permanent change in the behavior of electricity consumers will not be a better solution than the most advanced energy management algorithm.

**Author Contributions:** Conceptualization, P.P. and P.S.; methodology, P.P.; software, P.P.; validation, P.P. and P.S.; formal analysis, P.P.; investigation, P.P.; resources, P.P. and P.S.; data curation, P.P.; writing—original draft preparation, P.P. and P.S.; writing—review and editing, P.P. and P.S.; visualization, P.P.; supervision, P.P.; project administration, P.P. and P.S.; funding acquisition, P.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** This research was partially supported by funds allocated by the Rector of the University of Zielona Góra for the purchase of equipment necessary to conduct simulation studies.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

ADPA	adaptive dynamic programming algorithm
‘cheap’, ‘normal’, ‘expensive’	levels representing situations what the electricity price is
cost	includes the electricity price for a given tariff, control signals from the DSO resulting from the need to counteract peak demand, and weather conditions
DR	demand response
DSO	distribution network operators
EC	electrical appliance
EEM	elastic energy management
energy price	input data (calculation) allowing for the determination of the cost at the time specified by time based on the data presented in Figure 3
HEMS	home energy management system
‘increase consumption’, ‘reduce consumption’, ‘do not modify consumption’	defined messages that will be presented to the user to help them decide on modifying the EC operation
JSON	JavaScript Object Notation
kitchen hood, microwave oven, etc.	input data (observation) indicating the power consumption values of individual EC
message	response for input data considering Algorithm 1
ml	machine learning models
$P_{HEMS_H}$	the upper power threshold
$P_{HEMS_L}$	the lower power threshold
power sum	input data (calculation) representing the total power consumption of all EC at the time specified by time
$P_{SA}$	power consumption of smart appliance
RES	renewable energy sources
RL	reinforcement learning
SA	smart appliances
sc	test scenario
SVM	support vector machine
time	input data (observation) indicating the time at which the event occurred



## References

1. Ciesielka, E.; Dybowski, P.; Wójcik, J. Changes in the energy market structure and their impact on the formation of tariffs for electricity sales and distribution services. Optimization of purchase costs. *Przeegląd Elektrotechniczny* **2024**, *8*, 26–29. [CrossRef]
2. Siano, P. Demand response and smart grids—A survey. *Renew. Sustain. Energy Rev.* **2014**, *30*, 461–478. [CrossRef]
3. Nasab, M.A.; Hatami, M.; Zand, M.; Nasab, M.A.; Padmanaban, S. Demand side management programs in smart grid through cloud computing. *Renew. Energy Focus* **2024**, *51*, 100639. [CrossRef]
4. Benysek, G.; Jarnut, M.; Werminski, S.; Bojarski, J. Distributed active demand response system for peak power reduction through load shifting. *Bull. Pol. Acad. Sci. Tech. Sci.* **2016**, *64*, 925–936. [CrossRef]
5. Hadi, M.B.; Moeini-Aghtaie, M.; Khoshjahan, M.; Dehghanian, P. A Comprehensive Review on Power System Flexibility: Concept, Services, and Products. *IEEE Access* **2022**, *10*, 99257–99267. [CrossRef]
6. Dutta, G.; Mitra, K. A literature review on dynamic pricing of electricity. *J. Oper. Res. Soc.* **2017**, *68*, 1131–1145. [CrossRef]
7. Davarzani, S.; Pisica, I.; Taylor, G.A.; Munisami, K.J. Residential Demand Response Strategies and Applications in Active Distribution Network Management. *Renew. Sustain. Energy Rev.* **2021**, *138*, 110567. [CrossRef]
8. Liu, Y.; Xiao, L.; Yao, G.; Bu, S. Pricing-Based Demand Response for a Smart Home With Various Types of Household Appliances Considering Customer Satisfaction. *IEEE Access* **2019**, *7*, 86463–86472. [CrossRef]
9. Creutzig, F.; Niamir, L.; Bai, X.; Callaghan, M.; Cullen, J.; Díaz-José, J.; Figueroa, M.; Grubler, A.; Lamb, W.F.; Leip, A.; et al. Demand-side solutions to climate change mitigation consistent with high levels of well-being. *Nat. Clim. Chang.* **2022**, *12*, 36–46. [CrossRef]
10. Hao, C.H.; Wesseh, P.K.; Wang, J.; Abudu, H.; Dogah, K.E.; Okorie, D.I.; Osei Opoku, E.E. Dynamic pricing in consumer-centric electricity markets: A systematic review and thematic analysis. *Energy Strategy Rev.* **2024**, *52*, 101349. [CrossRef]
11. Freier, J.; von Loessl, V. Dynamic electricity tariffs: Designing reasonable pricing schemes for private households. *Energy Econ.* **2022**, *112*, 106146. [CrossRef]
12. Miletić, M.; Gržanić, M.; Pavić, I.; Pandžić, H.; Capuder, T. The effects of household automation and dynamic electricity pricing on consumers and suppliers. *Sustainable Energy, Grids and Networks* **2022**, *32*, 100931. [CrossRef]
13. Li, Z.; Wu, L.; Xu, Y.; Zheng, X. Stochastic-Weighted Robust Optimization Based Bilyer Operation of a Multi-Energy Building Microgrid Considering Practical Thermal Loads and Battery Degradation. *IEEE Trans. Sustain. Energy* **2022**, *13*, 668–682. [CrossRef]
14. Kim, B.G.; Zhang, Y.; van der Schaar, M.; Lee, J.W. Dynamic Pricing and Energy Consumption Scheduling with Reinforcement Learning. *IEEE Trans. Smart Grid* **2016**, *7*, 2187–2198. [CrossRef]
15. Moghaddam, V.; Yazdani, A.; Wang, H.; Parlevliet, D.; Shahnian, F. An Online Reinforcement Learning Approach for Dynamic Pricing of Electric Vehicle Charging Stations. *IEEE Access* **2020**, *8*, 130305–130313. [CrossRef]
16. Fraija, A.; Henao, N.; Agbossou, K.; Kelouwani, S.; Fournier, M.; Nagarsheth, S.H. Deep reinforcement learning based dynamic pricing for demand response considering market and supply constraints. *Smart Energy* **2024**, *14*, 100139. [CrossRef]
17. Sudha, N.; Santhiya, M.; Tamil Selvi, S.; Sathish, R. HEMS using Machine Learning and IoT in Energy Management. *Int. Res. J. Adv. Eng. Hub (IRJAEH)* **2024**, *2*, 1652–1658. [CrossRef]
18. Zhao, X.; Gao, W.; Qian, F.; Ge, J. Electricity cost comparison of dynamic pricing model based on load forecasting in home energy management system. *Energy* **2021**, *229*, 120538. [CrossRef]
19. Wan, Z.; Huang, Y.; Wu, L.; Liu, C. ADPA Optimization for Real-Time Energy Management Using Deep Learning. *Energies* **2024**, *17*, 4821. [CrossRef]
20. Feng, C.; Shao, L.; Wang, J.; Zhang, Y.; Wen, F. Short-term Load Forecasting of Distribution Transformer Supply Zones Based on Federated Model-Agnostic Meta Learning. *IEEE Trans. Power Syst.* **2024**, 1–13. [CrossRef]
21. Shavolkin, O.; Shvedchikova, I.; Kolcun, M.; Medved, D.; Mazur, D.; Kwiatkowski, B. Increasing photovoltaic self-consumption for objects using domestic hot water systems. *Arch. Electr. Eng.* **2024**, *73*, 573–593. [CrossRef]
22. Unhelkar, B.; Pandey, H.; Agrawal, A.; Choudhary, A. *Advances and Applications of Artificial Intelligence & Machine Learning: Proceedings of ICAAAIML 2022; Lecture Notes in Electrical Engineering*; Springer: Singapore, 2023.
23. Powroźnik, P.; Szcześniak, P.; Sobolewski, Ł.; Piotrowski, K. Novel Functionalities of Smart Home Devices for the Elastic Energy Management Algorithm. *Energies* **2022**, *15*, 8632. [CrossRef]
24. Powroźnik, P.; Szcześniak, P.; Piotrowski, K. Elastic Energy Management Algorithm Using IoT Technology for Devices with Smart Appliance Functionality for Applications in Smart-Grid. *Energies* **2022**, *15*, 109. [CrossRef]
25. Energy Consumption Meter SEM8500 6-socket Wi-F. Available online: <https://www.conrad.com/en/p/voltcraft-sem8500-energy-consumption-meter-wireless-sockets-data-export-mode-data-logger-energy-cost-calculator-se-2359015.html> (accessed on 24 October 2024).
26. Tuya—A Global Cloud Platform That Enables The Creation And Management Of Smart Devices. Available online: <https://www.tuya.com/> (accessed on 24 October 2024).

27. Marrs, T. *JSON at Work: Practical Data Integration for the Web*; O'Reilly Media: Sebastopol, CA, USA, 2017.
28. Pachouly, J.; Ahirrao, S.; Kotecha, K.; Selvachandran, G.; Abraham, A. A systematic literature review on software defect prediction using artificial intelligence: Datasets, Data Validation Methods, Approaches, and Tools. *Eng. Appl. Artif. Intell.* **2022**, *111*, 104773. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.