

Short-Term Load Forecasting on Individual Consumers

João Victor Jales Melo ^{1,*}, George Rossany Soares Lira ², Edson Guedes Costa ², Antonio F. Leite Neto ¹
and Iago B. Oliveira ¹

¹ Postgraduate Program in Electrical Engineering, Federal University of Campina Grande, Campina Grande 58428-830, Brazil

² Electrical Engineering Department, Federal University of Campina Grande, Campina Grande 58428-830, Brazil

* Correspondence: joao.melo@ee.ufcg.edu.br

Abstract: Maintaining stability and control over the electric system requires increasing information about the consumers' profiling due to changes in the form of electricity generation and consumption. To overcome this trouble, short-term load forecasting (STLF) on individual consumers gained importance in the last years. Nonetheless, predicting the profile of an individual consumer is a difficult task. The main challenge lies in the uncertainty related to the individual consumption profile, which increases forecasting errors. Thus, this paper aims to implement a load predictive model focused on individual consumers taking into account its randomness. For this purpose, a methodology is proposed to determine and select predictive features for individual STLF. The load forecasting of an individual consumer is simulated based on the four main machine learning techniques used in the literature. A 2.73% reduction in the forecast error is obtained after the correct selection of the predictive features. Compared to the baseline model (persistent forecasting method), the error is reduced by up to 19.8%. Among the techniques analyzed, support vector regression (SVR) showed the smallest errors (8.88% and 9.31%).



Citation: Melo, J.V.J.; Lira, G.R.S.; Costa, E.G.; Leite Neto, A.F.; Oliveira, I.B. Short-Term Load Forecasting on Individual Consumers. *Energies* **2022**, *15*, 5856. <https://doi.org/10.3390/en15165856>

Academic Editors: Sergio Gómez Melgar, Daniel Villanueva Torres and Ali Hainoun

Received: 9 March 2022

Accepted: 19 April 2022

Published: 12 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

Keywords: load forecasting; machine learning; neural network; smart meter

1. Introduction

The electrical system is undergoing a modernization process. The new ways of generating and consuming energy increasingly increase the complexity of its control and operation. The growing use of renewable sources such as wind and solar are examples of this process. These sources are characterized by not allowing control on demand, being intermittent and difficult to predict [1]. The share of renewable energy sources in power generation shall reach up to 65% in 2050 [2]. The system's complexity grows in parallel with this process. In this context, information on the load side and predictions of its behavior are essential for the safety and stability of the system.

The ability to measure and interpret data in real-time at each consumer unit makes it possible to obtain new services such as estimating the daily consumption curve and controlling the load of each individual consumer. The smart meters are presented as the connection point for creating and distributing this information. Unlike conventional meters, data analysis, load management, and load forecasting can be achieved using smart meters [3,4]. This new information allows the consumer himself to contribute to more rational use of energy.

Thus, the individual short-term load forecasting (STLF) in smart meters gains importance as part of the progress towards a more interconnected system. Individual short-term load forecasting can help the demand side management, (DSM) [1,5] and the response demand (DR) [6], which creates more stability in the system through indirect control of electricity consumption.

Improving load forecasting is the goal of several studies [4–18]. However, studies on individual consumers are scarce. The reason for this is that the low or medium voltage load forecast has a high degree of randomness and dependence on the characteristics of each consumer, which reduces the accuracy of the forecasts.

The papers that perform STLF use different methodologies to reduce their errors. One of them is to aggregate the measurement of several smart meters, making the prediction of the aggregate load [1,7]. A second way found is to group measurements from consumers with similar behavior to train machine learning algorithms [4,6,8]. With these methodologies it is possible to reduce the randomness of the database, however, it creates a dependency on the expected characteristic for the data and an infrastructure of meters is needed. Another proposed methodology consists of pre-processing the database, which can be done by treating the noise in the database [9] or by selecting the best features for prediction through feature engineering [10].

Treating the database and selecting the best features can significantly reduce the prediction error. Therefore, the present work aims to develop a methodology for short-term load forecasting in individual consumers. The methodology consists of using feature engineering and feature extraction techniques to select the most significant features from the database and, consequently, reduce the forecast error associated with the randomness of the measurements.

The main contributions of this paper can be summarized as follows:

- Propose the use of the mutual information technique as one of the criteria to select the most important features for load forecasting;
- Propose a methodology for choosing significant features based on the joint application of mutual information and correlation techniques;
- Compare STLF results with different machine learning techniques;
- Improved short-term load forecasting in individual consumers by pre-processing and feature selection.

Predictive models of electricity consumption will be developed in association with the main machine learning techniques found in the literature with application in short-term load forecasting.

The four machine learning techniques implemented are:

- Support Vector Regression (SVR);
- Feedforward Multilayer Perceptron (MLP);
- Long short-term memory (LSTM);
- ARTMAP-Fuzzy.

The forecasting models and the methodology for selecting the features are evaluated for error in tests performed with a database of an individual consumer smart meter belonging to the distribution system of Paraíba in Brazil.

The rest of this paper is organized as follows: in Section 2 we review the relevant literature. Section 3 describes our methodology for short-term load forecasting and Section 4 presents the results of our tests on an actual load database. In Section 5 we discuss our results and provide an outlook for future research.

2. Related Work

The energy consumption of a user is composed of the set of devices that demand electrical energy and that together create a consumption pattern represented by a load curve. Load forecasting can be performed from historical consumption data and other attributes that influence energy consumption. Many works use this data with machine learning techniques to predict the behavior pattern of consumers. In this section, some of the recent approaches related to STLF are briefly presented. The papers are summarized in Table 1.

Table 1. Summary of the related works.

Reference	Techniques	Granularity
[11]	LSTM	30 min
[12]	LSTM	30 min
[13]	ARTMAP-Fuzzy	15 min
[14]	K-medoids and ANN	Daily
[10]	Sliding window-based LightGBM	15 min
[15]	Stacking ensemble MLP	15 min
[16]	RF, GB and SVM, GRU and CNN	Daily
[17]	BSTS	1 h
[18]	GRU	1 h
[19]	Non-Iterative Ito Decomposition and SGTm	-
[20]	Non-iterative SGTm	-
[21]	State Estimation NN	30 min

Kong et al. [11] proposed the use of an LSTM neural network as a way of identifying the consumption pattern and making a 30-min load forecasting for a residential unit. They used a vast database with 3 months of consumption of 69 consumer units. The load curve of each residence is predicted individually. The results are compared with the state of art techniques, including the MLP neural network and the K-Nearest Neighbor (KNN) algorithm. The LSTM was considerably better than the others analyzed, with an absolute mean percentage error (*MAPE*) of 44.06%, while the best result among the others analyzed was 49.49%.

Kong et al. [12], in another paper, used a network of sensors spread over a residence, measuring the consumption of the main loads to help predict the total load. The LSTM neural network with a forecast horizon of 30 min is used. The work achieves a considerable reduction in forecast error when compared to its previous work. The results of the proposed model are compared with MLP and KNN networks. The LSTM network is considerably better than the others analyzed.

Alves [13] used an ARTMAP-Fuzzy neural network for STLF for individual users. The maximum, minimum, and average values of energy consumption are extracted and used as input for the predictor model in addition to the time series of energy consumption and temporal variables. A considerable improvement in the prediction of energy consumption is observed when the new extracted attributes are used. The *MAPE* of forecasts ranged from 17.06% to 38.09%.

Similar to [12], Haq et al. [14] performed the forecast of the daily demand peak of a residence from the measurement in its main loads. The authors carry out the consumption forecast of each load individually and consider the results as inputs for the forecast of the daily demand peak, obtaining a considerable reduction of the forecast error.

In Park et al. [10], authors use the pre-processing of data with identification and correction of errors and the creation and selection of features to improve the result of the prediction model. Pearson's correlation coefficient is used as a selection criterion for predictive features. Temporal and climatic variables and statistical data extracted from these time series are taken into account during the feature extraction process. The work used a sliding window in time to simulate an online forecast. As a result, the relationship between the prediction improvement and the computational cost with the window growth is observed.

Moon et al. [15] carried out a similar work focused on load of a residential building. The authors study the optimal data window for carrying out the short-term load forecast and use Pearson's coefficient as a criterion for identifying the most significant attributes.

The proposed prediction technique is called COSMOS, based on the stacking ensemble of MLP neural networks. The authors also demonstrate the importance of optimizing the hyperparameters of the forecasting technique, which has a significant impact on the final error.

In Ayub et al. [16] feature extraction and selection are used to improve short-term load forecasting. The authors propose the use of Random Forest and Extreme Gradient Boosting techniques to determine the importance of features. In addition, the RFE technique is used to determine the redundant features, removing the less important features for the forecast. As a result, the article achieves an accuracy of up to 96.33% in the STLF.

Roth et al. [17] use a database of 120 smart meters of residential apartments in Singapore and forecast hourly consumption at the individual level. The apartments have smart meters in the main electrical equipment and these data are considered during the forecast. To prepare the database, a feature selection process is carried out using the Spike-and-Slab method presented in [22], in which the most significant features for prediction are selected. As a result, an improvement in the forecast is observed with the use of features related to measurements in the main loads of the residence, mainly with larger forecast horizons.

In Jung et al. [18] the STLF model is proposed using attention-based GRU neural network to give more weight to the features with more relevance for forecasting. Through tests carried out with data from real consumers, an improvement in the prediction is proven when compared to state-of-the-art techniques.

As demonstrated above, many works use interactive techniques to perform the STLF, as is the case with [15,16]. Due to the characteristics of these methods, they are associated with long training times and their results are difficult to interpret. Alternatively, the author of [23] describes a new topology and training algorithm for a non-interactive neural structure. The approach is based on a Successive Geometric Transformations Model (SGTM) neural link. With the new topology, advantages such as faster training, data pre-processing, selection of the most significant attributes, and flexibility in solving different types of tasks are achieved. Some authors have applied the topology to prediction tasks [19] or for optimized identification of regression coefficients [20], demonstrating the feasibility of the technique.

Prakash and Sydulu [21] develop a similar methodology with the aim of performing the STLF in a non-interactive way. For this, a state estimation approach is used to determine the ideal set of weights of neurons in a neural network. The methodology is applied to real load data in a power distribution system. As a result, STLF is achieved with an error like a common neural network, but with a drastic reduction in training time.

As can be confirmed by the review presented, the challenges encountered in load forecasting of individual users are still present and are the motivation for the most recent work in the STLF. To improve load forecasting, some works make use of measurement at several points as in [12,14,17]. These works have the problem of needing a structure of meters spread throughout the residence. Other works like [15,16,21,23] use pre-processing and feature selection techniques to improve load forecasting. These works demonstrate that the correct selection and treatment of data is capable of significantly reducing the forecast error.

In this paper, the author develops a system for STLF of individual consumers, using data pre-processing techniques and the selection of the best features to reduce errors during the forecast. As a differential to the other works, the author proposes a method to select the input set. Unlike some works that use correlation as a mathematical criterion for selecting features [10,15], it is proposed to use a combination of correlation calculation and mutual information. The use of mutual information allows the analysis of linear and non-linear dependence between variables. In this way, the selection is performed by mathematical criteria, avoiding interactive methods [15,16] and considering all the dependency relationships between the features of the database. In addition, the main features for load forecasting are extracted by a feature engineering process based on the features proposed in the literature review.

3. Methodology

The proposed methodology consists of using feature engineering and extraction techniques to select the most significant features from the database. The main machine learning techniques are used to perform modeling and load forecasting.

Figure 1 shows the steps proposed to carry out this work.

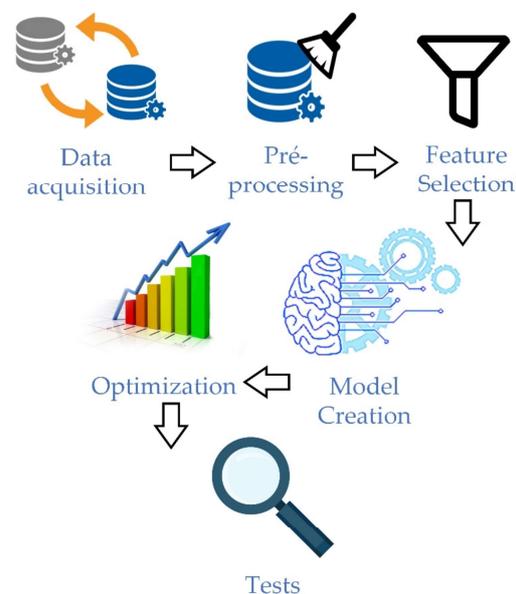


Figure 1. Flowchart of the methodology used.

As shown in Figure 1, the methodology implemented is composed of six sequential steps that aim to perform data processing, extract information, and select the most explanatory predictor features to forecast electricity consumption. Each step of the methodology will be described in detail below, as well as the tools used for its implementation.

3.1. Data Acquisition and Pre-Processing

The database used belongs to an individual consumer in the distribution network of the state of Paraíba, Brazil.

The analyzed consumer analyzed has a differentiated time-of-use tariff and is characterized by a maximum demand of up to 43 kW. A set of 1522 measurements performed every 30 min from 28 March 2019 to 29 April 2019 composes the database. The variables are presented in Table 2.

Table 2. Database variables.

Type	Feature
Predictors	Energy Consumption (t–25 to t–1) ¹
	Day (t)
	Hour (t)
	Rush hour (t)
Objective	Energy Consumption (t)

¹ "t–n" is the number of measurements prior to the current value "t".

As shown in Table 2, the database is formed by a set of predictor variables and the objective variable. The predictor variables are composed of the time series of energy consumption measured every 30 min, hour, day, and rush hour. The objective variable, in turn, is the energy consumption to be forecast 30 min ahead.

Once the database is formed, a pre-processing step is performed, in which missing or possibly erroneous data is checked, and their correction is made. For the load time

series, the wrong or missing data is replaced by an average of the consumption measured in the time before and after the error, as shown in (1). For errors in temporal variables, the correction is made from the information in the previous time interval.

$$C_t = C_{t-1} - \frac{C_{t+1} - C_{t-1}}{2}, \quad (1)$$

where C_t is the corrected load, C_{t-1} is the load in the previous period and C_{t+1} is the load in the period after the error.

Still, during pre-processing, all variables are normalized in order to prevent the forecasting algorithm from becoming biased towards the variables of the highest order of magnitude. In this way, all information is transformed to be in the range of 0 and 1. In (2) the equation used for normalization is presented.

$$X_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (2)$$

where X_{norm} is the normalized variable, x is its actual value, x_{max} is the maximum value of the time series before normalization and x_{min} is the minimum value of the time series before normalization.

3.2. Feature Selection

In the feature selection stage, the database is analyzed to determine the information that most influences an accurate forecast of electricity consumption. For this, the database's own information is used in addition to extracting new features that have the explanatory capacity for the forecasting task. This data mining process is called feature engineering and can be done by associating information present in the database or by mathematical relationships, as is the case with maximum and minimum daily consumption values, bringing new information to the predictor algorithm.

With a determined set of predictive features, the optimum set that maximizes the accuracy of the forecast must be selected. The tasks of extracting information and selecting the forecast set are performed following the steps indicated in the flowchart of Figure 2.

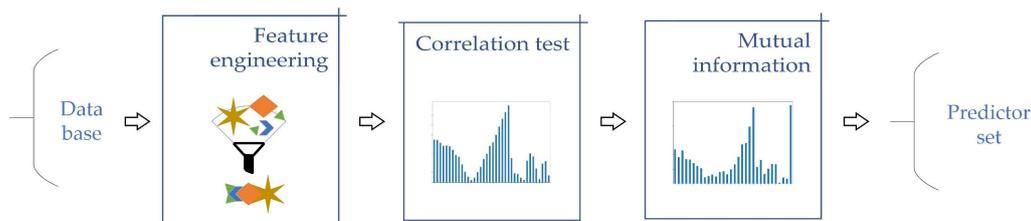


Figure 2. Proposed solution for extracting features and selecting the most important ones.

As shown in Figure 2, the first step in determining the forecast set is to analyze the database with the objective of extracting as much information as possible to assist in the prediction of electricity consumption. For this purpose, the features presented in Table 3 are calculated, which statistically evaluate the time series of energy consumption in terms of position, dispersion, and distribution.

Table 3. Evaluated features.

Type of Feature	Feature
Position	Maximum
	Minimum
	Average
	Median
	Quantile (15%) ¹
Dispersion	Range
	Variance
	Standard deviation
Distribution	Obliquity
	Kurtosis

¹ Value that is not exceeded in 15% of measurements.

As a short-term load forecast is sought, only the set of data measured in the last 12 h is considered for the calculation of the features in Table 3. Thus, only short-term information is attributed to these features. The new information is added to the database for further analysis.

The next step in selecting the features is to determine which information has the greatest explanatory capacity for forecasting energy consumption. This is accomplished by analyzing the correlation and mutual information (MI) of the features in the database.

The correlation between the features of the database and the energy consumption forecast is calculated with the objective of discarding features that have a weak correlation, which possibly indicates that the information will not be relevant to the forecast model or, even, will harm the results.

Finally, the mutual information (MI) analysis is performed among the remaining features. Mutual information is a measure of similarity that determines the amount of information that a variable contains about another and that, unlike correlation, could identify both linear and non-linear relationships [24]. Thus, the features are analyzed concerning MI to identify redundancies in the database, keeping only those that have a higher MI with the objective variable. The criterion adopted for a pair of features to be considered redundant is to be greater than 0.6, thus, more than half of the information that the variables have is the same.

Thus, after the extraction and analysis of the features, the forecast set chosen must be formed by variables loosely linked to each other, that is, that individually bring new information to the forecast system, and that have a high explanatory capacity in relation to the forecast.

3.3. Model Creation

After the procedures described above, the database formed can be submitted to a model based on machine learning to perform the load forecast.

Four forecasting models using the main forecasting techniques found in the scientific bibliography were implemented. These are:

- Support vector regression (SVR);
- Feedforward multilayer perceptron (MLP);
- Long short-term memory (LSTM);
- ARTMAP-Fuzzy.

Forecasting models were created with a 30-min and 6-h horizon (with a 30-min step between predictions). Due to the differences found between the techniques, the models created with the MLP and LSTM networks differ from the others during the forecast with a 6-h horizon. Figures 3 and 4 show the ways chosen to use the forecast models according to their limitations.

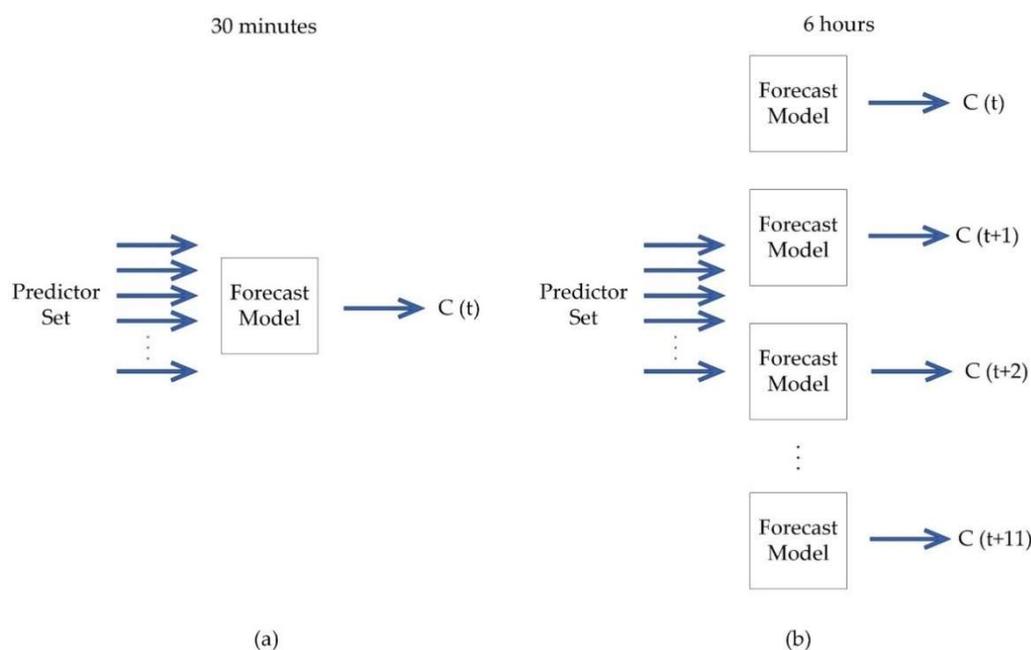


Figure 3. Predictive models for SVR and ARTMAP-Fuzzy network with a horizon of (a) 30 min and (b) 6 h.

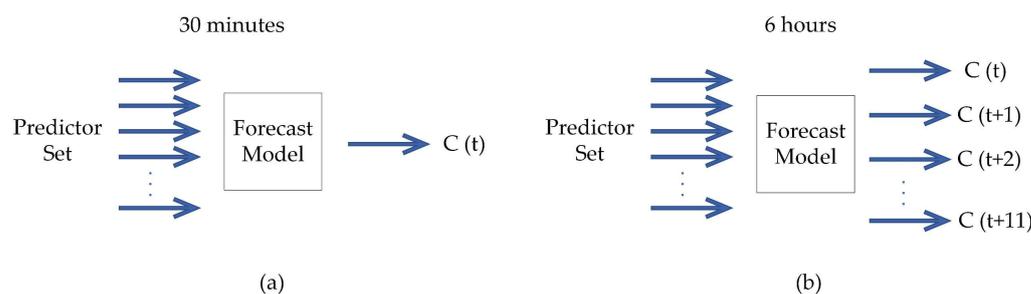


Figure 4. Predictive models for MLP and LSTM networks with a horizon of (a) 30 min and (b) 6 h.

As can be seen in Figures 3 and 4, when using SVR or the ARTMAP-Fuzzy network, a model is trained for each output. In another way, for the MLP and LSTM neural networks, which have the output formed by neurons, only one model can be trained with the number of neurons required for the desired output, as shown in Figure 4.

The database was separated into two sets (75% and 25%) for each predictor system, where 75% of the data is used for training and 25% is used for testing and validation.

3.4. Optimization

The models described in the previous subsection require the definition of hyperparameters during their implementation. Hyperparameters quantify properties of predictive systems that are constant throughout the database samples and that condition and establish non-linear relationships with the other parameters [25]. In other words, it can be said that hyperparameters are the initial configurations given to the forecasting system, defined from the previously known characteristics of the database.

A good choice of hyperparameters has a great influence on the performance of the predictive model. For this, there are several search techniques with the function of determining the optimal set for a model. One that can be implemented and that is suitable for searches with few variables is the randomized optimization of parameters [26].

From a defined set of hyperparameters and a specified search range, the randomized parameter optimization method (RPO) tests the predictor system by varying the hyperparameters within a probabilistic distribution, until the most accurate models are found.

The RPO is used to determine the best set of hyperparameters for the four models described in the previous subsection. Table 4 shows the defined hyperparameters to be optimized, the search limit and the number of interactions performed.

Table 4. Hyperparameter range used in optimization.

Forecast Model	Hyperparameter	Bottom Limit	Upper Limit	N°
SVR	Epsilon	0.001	0.01	50
	C	0	100	
MLP	N° Layers	1	5	60
	N° Neurons	5	20	
LSTM	N° Layers	1	5	72
	N° Memory Loops	7	25	
ARTMAP-Fuzzy	Surveillance	0.8	0.98	200
	Alpha	0.0001	0.01	

The choice of the search limits takes into account the optimal value expected for each hyperparameter according to the typical values found in the literature. Some hyperparameters were set with typical values used for time series forecasting, as shown in Table 5 [27,28].

Table 5. Non-optimized hyperparameters.

Forecast Model	Hyperparameter	Value
SVR	Kernel	RBF
MLP	Activation	Sigmoid
	Training	Backpropagation
LSTM	Activation	Hyperbolic Tangent
	Recurring Activation	Sigmoid
ARTMAP-Fuzzy	Training Rate	1
	Tracking	0.001

With the definition of hyperparameters, the predictive models are ready to be trained and used.

3.5. Tests

The evaluation of a regressive predictor model requires the use of metrics that verify the robustness of the solution, indicating how close the model is to reality. Essentially, the performance metric must be taken into account the individual error of each forecast and the model's deviation from real values. For this, there are several metrics consolidated in statistics and data science. Some of the main ones are the mean absolute percentage error (*MAPE*), which offers a measure of the error of all estimates made by the model, and the coefficient of determination (R^2), which indicates how fitted the model is in relation to the real values.

MAPE is a measure of the accuracy of a model that considers each estimate error. The *MAPE* result ranges from 0 to 100, indicating that the model is more accurate the lower its value. The calculation is defined, as the name implies, by the average percentage error of all estimates, as shown in (3).

$$MAPE = \frac{1}{n} \sum_{T=1}^N \left| \frac{\hat{y}_i - y_i}{\hat{y}_i} \right|, \quad (3)$$

where y_i are the model estimates, \hat{y}_i are the real values and n is the size of the vector y_i .

The coefficient of determination (R^2), in turn, is a metric for determining the fit of a model to real values. Its result ranges from 0 to 1 and indicates that the model is more accurate the higher its result. Its calculation is performed as shown in (4).

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4)$$

where, y_i are the model estimates, \hat{y}_i are the real values and \bar{y} is the average value of all measurements.

The error assessment metrics implemented in this work will be the *MAPE* and the R^2 coefficient. The objective is to determine the model that presents the best results in relation to the forecast accuracy.

4. Results

In this section are presented the results of the feature selection, hyperparameter optimization, and the forecast error of the developed models for the short-term load forecasting of an individual consumer.

After presenting the results of the forecasting models designed from the proposed methodology, the forecasting errors will be compared between the models based on selected features or not, identifying the contribution of this step to the accuracy of the models.

4.1. Feature Selection

As described in the methodology, the features in Table 3 are calculated with the objective of adding new information to the database to assist in the prediction of energy consumption. Uniting the calculated features with the features of the database, the set of variables presented in Table 6 is obtained.

Table 6. Set of evaluated features.

Feature
Energy Consumption (t–25 to t–1)
Day (t)
Hour (t)
Rush hour (t)
Maximum
Minimum
Average
Median
Quantile (15%)
Range
Variance
Standard deviation
Obliquity
Kurtosis

The correlation between each predictive feature indicated in Table 6 was calculated in relation to the objective variable (consumption (t)). The result is shown in Figure 5.

From the analysis of the graph in Figure 5, it can be seen that some features have a low correlation with the energy consumption forecast, especially C (t–14), C (t–13), C (t–12), median, quantile, and kurtosis. These variables were discarded because it was considered that they would not bring new information to the predictor model.

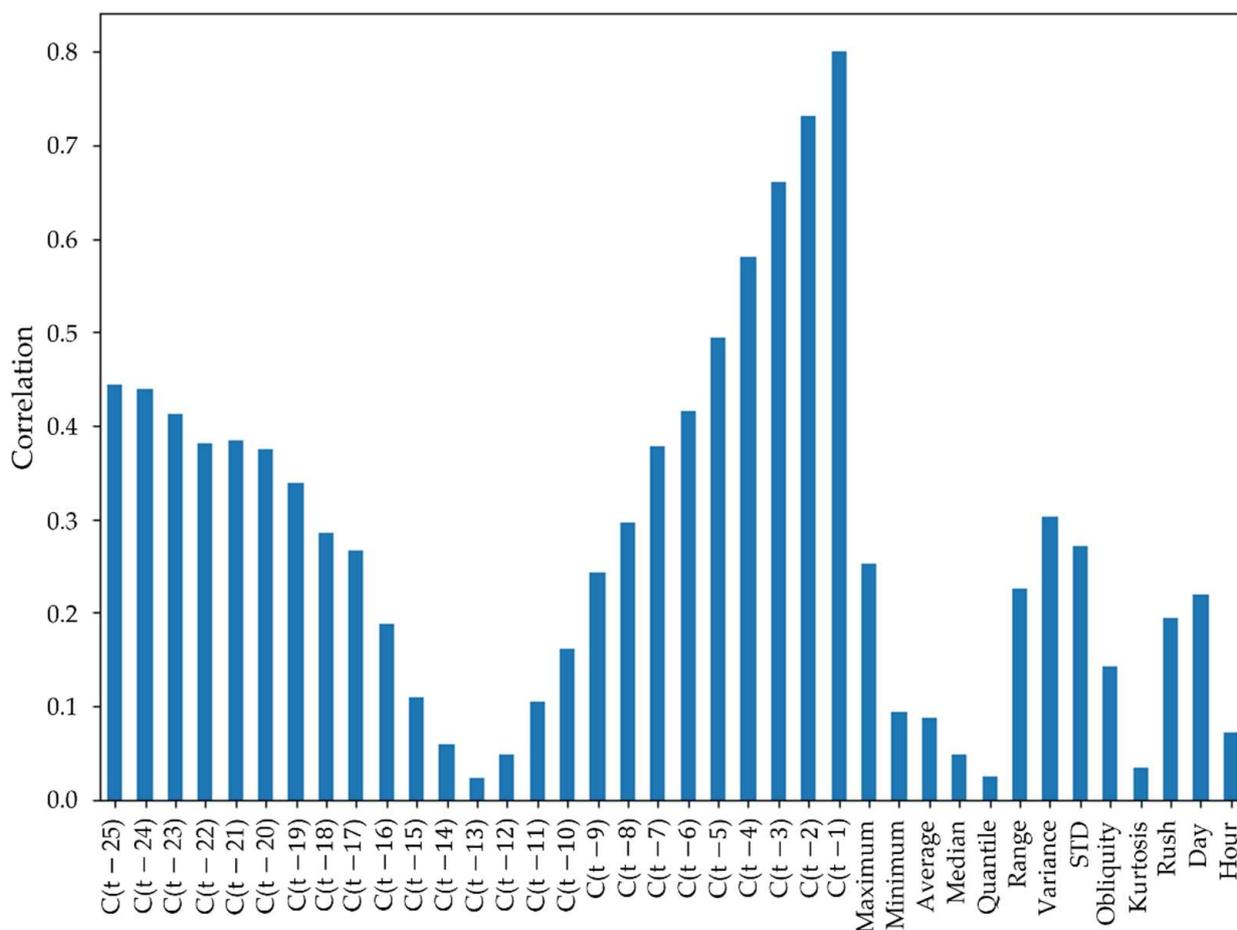


Figure 5. Bar graph of the correction of each feature with the forecast of energy consumption.

In order to determine the best set of features without redundancy in the input, the MI was calculated among all the remaining features in the database, thus determining the MI for each pair of remaining attributes. The result of this calculation is shown in Figure 6.

As can be seen, a strong relationship was found between the features presented in Table 7, in which 0.6 normalized MI, as proposed in the methodology of this paper, was exceeded.

Table 7. Features with MI greater than 0.6 between themselves.

Feature 1	Feature 2	MI Normalized
STD	Variance	0.9724
Rush hour	Hour	0.7849
Day	Maximum	0.7475
Maximum	Range	0.6303

In order to avoid redundancies in the set of inputs of the predictive model, the feature that presented the lowest MI in relation to the electricity consumption was discarded in each pair of variables in Table 7. The calculated values for the MI between the attributes of Table 7 and the electricity consumption are presented in Table 8.

As can be seen in Table 8, the “Variance”, “Hour” and “STD” have the highest MI values. As “Variance” and “STD” were classified as redundant attributes by the analysis of Table 7, the attribute with the lowest MI with energy consumption was discarded. Thus, the “Standard Deviation”, “Rush”, and “Range” were discarded from the features set.

The set of predictive features obtained at the end of the analysis is shown in Table 9.

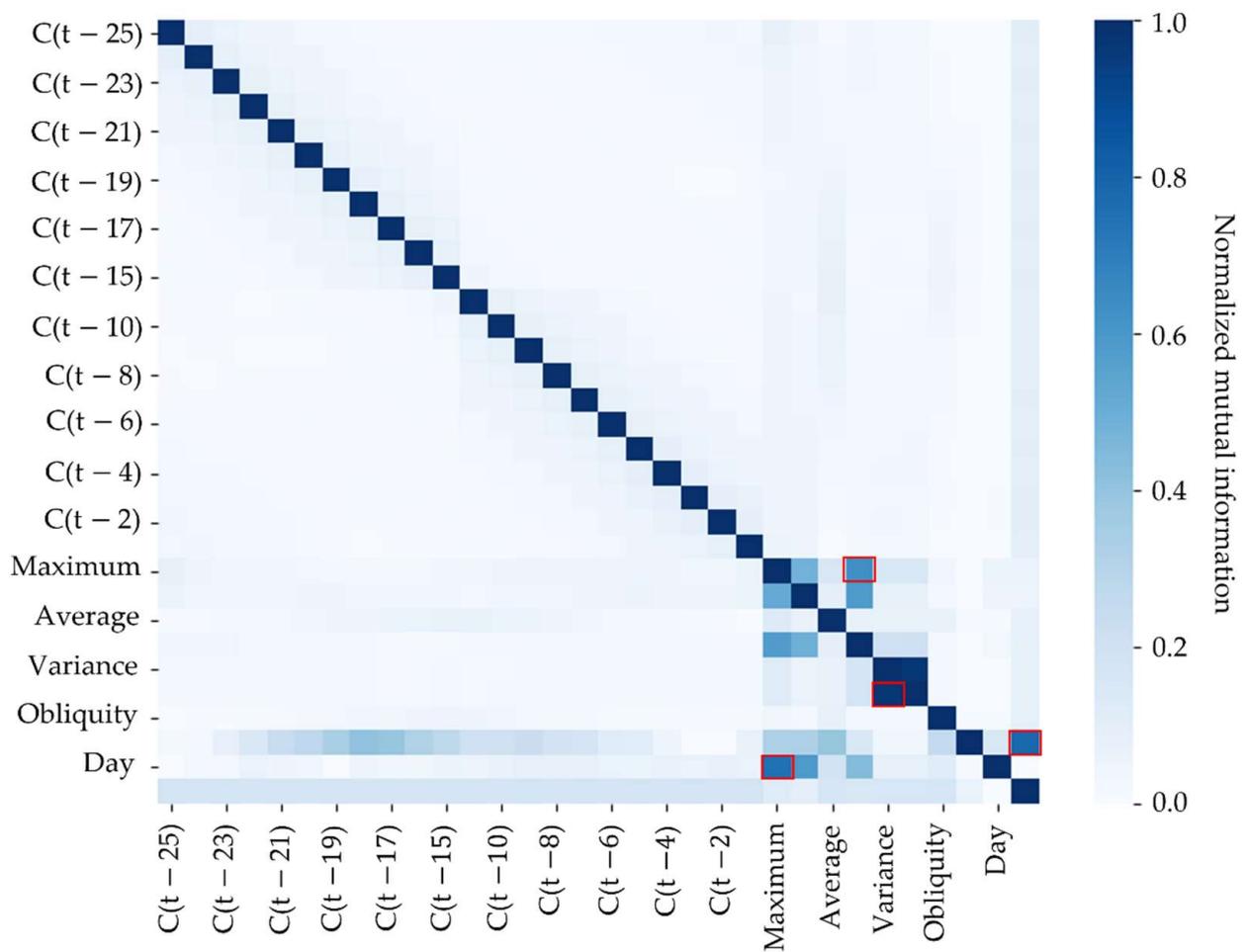


Figure 6. Normalized mutual information to each pair of features.

Table 8. Normalized mutual information of each feature regarding energy consumption.

Feature	MI Normalized
Hour	0.5480
Variance	0.1343
STD	0.1329
Maximum	0.1164
Range	0.1046
Rush	0.0286

Table 9. Selected feature set.

Feature
Energy Consumption (t–25 to t–15)
Energy Consumption (t–11 to t–1)
Day (t)
Hour (t)
Maximum
Minimum
Average
Variance
Obliquity

4.2. Optimization

After the predictive attributes were chosen, the hyperparameters to be used were determined according to the proposed methodology. The results are shown in Table 10.

Table 10. Optimized hyperparameters.

Forecast Model	Hyperparameter	Value
SVR	Epsilon	0.0012
	C	0.1633
MLP	N° Layers	2
	N° Neurons	15
LSTM	N° Layers	3
	N° Memory Loops	25
ARTMAP-Fuzzy	Surveillance	0.8
	Alpha	0.0071

4.3. Tests

The predictive models implemented using the selected attributes and the optimized hyperparameters were tested in two-time horizons, 30 min and 6 h (with a 30-min step between predictions). As previously described, tests were performed on 25% of the database, forming 380 forecasts (7 days and 22 h). Models are trained and tested until the average of their results varies by less than 5%. The models are analyzed for *MAPE* and R^2 and their average results are presented in Tables 11 and 12.

Table 11. *MAPE* and R^2 for 30-min horizon forecasts.

Forecast Model	<i>MAPE</i> (%)	R^2
SVR	8.88	0.72
MLP	10.62	0.62
LSTM	9.62	0.68
ARTMAP-Fuzzy	12.23	0.45

Table 12. *MAPE* and R^2 for 6-h horizon forecasts.

Forecast Model	<i>MAPE</i> (%)	R^2
SVR	9.31	0.67
MLP	11.02	0.56
LSTM	10.58	0.61
ARTMAP-Fuzzy	12.74	0.31

The SVR model, among all the tested models, presented the best performance for the two analyzed time horizons. Figure 7 presents the performance of the SVM model on load forecast in 30-min time horizon.

The accuracy improvements analysis in the best model, provided by the use of extracting information from the database and selecting the most explanatory features, is shown in Table 13, with the following inputs:

- Features listed in Table 9, which were obtained from the analysis and selection of the best set of database features, as described in the methodology.
- Features listed in Table 6, no feature selection.

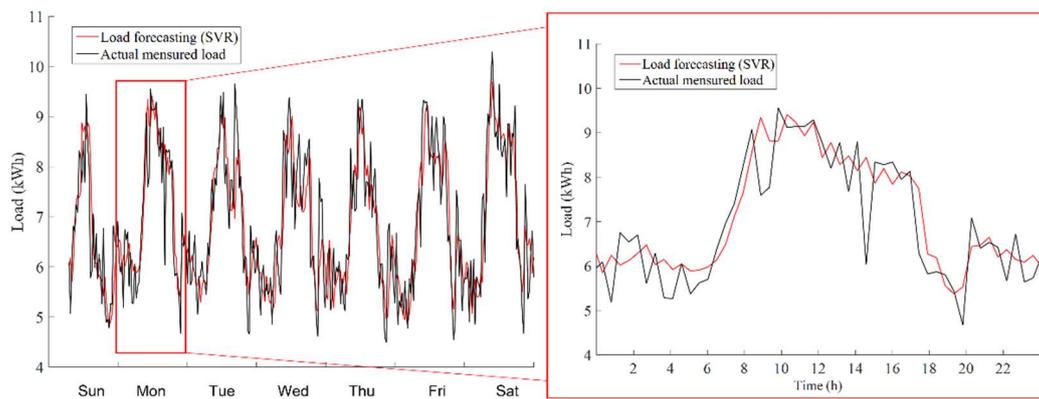


Figure 7. Load forecasting by the SVR model with 30-min time horizon.

Table 13. SVR *MAPE* and R^2 with both time horizons after feature selection.

Time Horizon	<i>MAPE</i> Reducing (%)	R^2 Increasing (%)
30 min	1.79	0.65
6 h	2.37	0.71

Although the results presented in Table 13 are the main analysis of this work, proving the effectiveness of using data mining and selection of variables, additional tests were performed to compare the performance of the predictors with a baseline approach. A persistent estimator was used as a point of comparison because it generally offers good results in short-term time series forecasts, especially those with wide variation [1]. The results are shown in Table 14.

Table 14. *MAPE* and R^2 with a 30-min horizon using the implemented models compared to a persistent model.

Forecast Model	<i>MAPE</i> Reducing to Baseline (%)	R^2 Increasing to Baseline (%)
SVR	19.80	36.31
MLP	4.08	17.38
LSTM	13.11	28.74
ARTMAP-Fuzzy	−1.05	−13.42

4.4. Discussion of Results

With the extraction and selection of attributes from the database, in a process of feature engineering and data mining, an explanatory predictive set for the consumption of electricity was achieved. As shown in Table 10, the dataset was applied in the SVR, MLP, LSTM, and ARTMAP-Fuzzy techniques, allowing the short-term load forecasting of an individual consumer with an error of up to 8.88%.

When analyzing the results presented in Tables 11 and 12, it can be seen that considering the performance metrics adopted, SVR was the best forecasting model followed by the LSTM network in the two horizons studied. As for the forecast error (*MAPE*), the SVR presented 8.88% and 9.31%, respectively for the horizon of 30 min and 6 h of the forecast. In relation to the R^2 metric, which indicates curve fitting quality, the SVR also obtained the best results, around 0.7, a very high value considering the randomness of the analyzed dataset.

As a point of comparison for the numerical results obtained, the persistent forecasting method (Table 14) with a forecast horizon of 30 min was adopted. Its results are used as a comparison because it is considered a hard-to-beat forecasting methodology in short-term horizons. As shown in Table 14, all forecasting models developed, with the exception of ARTMAP-Fuzzy, showed improvements in both the forecasting error (*MAPE*) and in

adjusting the curve (R^2), reaching 19.8% and 36.31% improvement, respectively, for *MAPE* and R^2 of the SVR model.

Finally, with the analysis of Table 13, an improvement of the predictive model is verified due to the correct extraction and selection of attributes. A relative improvement of 1.79% in the forecast error (*MAPE*) was achieved with a 30-min horizon and 2.37% with a 6-h horizon.

5. Conclusions

In this paper, a methodology was developed to extract and select the best set of attributes in short-term load forecasting. The extraction of attributes from a database was carried out from a statistical analysis taking into account the position, dispersion, and distribution of the time series. The selection process was carried out through correlation analysis and MI, indicating a set of explanatory and independent features. The methodology was applied together to 4 (four) of the well-known time series forecasting techniques in order to predict the consumption of an individual consumer.

The results obtained indicate an improvement in the short-term load forecasting of individual consumers. The forecast error (*MAPE*) decreased by 19.8% and the curve fitting (R^2) increased by 36.31% in the best case (SVR) when compared to the baseline approach.

Taking into account only the extraction and selection of attributes and keeping forecasting techniques constant, an improvement of 1.78% was achieved in the 30-min horizon and 2.37% in the 6-h horizon of load forecasting. In this way, it can be said that the selected dataset contributed to the improvement of the final models, making clear the contribution and importance of the treatment and selection of features from the database for a correct forecast of energy consumption.

The methodology implemented has the limitation of requiring the determination of limits of importance for the feature selection, both for correlation and for MI. This limitation can affect the accuracy of prediction models due to the need for specialized human intervention, which also makes scalability difficult. In future works, optimization techniques may be proposed that determine the best values without the need for intervention.

The prediction of very random loads, such as the individual consumer analyzed in this paper, is still emerging research, but it has great potential in the optimization of the electrical system. We consider that the results presented in this paper contribute to the development of the area and growth of the characterization of the consumer profile.

Author Contributions: Conceptualization, J.V.J.M. and G.R.S.L.; methodology, J.V.J.M.; software, J.V.J.M.; data curation, J.V.J.M.; supervision, G.R.S.L.; project administration, E.G.C.; writing—original draft preparation, J.V.J.M.; writing—review and editing, A.F.L.N., I.B.O. and G.R.S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), great number: 88887.641227/2021-00 and Universidade Federal de Campina Grande (UFCG), great number: 23096.003067/2022-01. The third author acknowledges the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the Research Productivity Grant (CNPq N°. 09/2020-PQ). The authors are grateful for the funding grant number 04/2022, Paraíba State Research Foundation (FAPESQ).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Ziekow, H.; Goebel, C.; Strucker, J.; Jacobsen, H. The potential of smart home sensors in forecasting household electricity demand. In Proceedings of the 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm), Vancouver, BC, Canada, 21–24 October 2013. [\[CrossRef\]](#)
2. Gielen, D.; Boshell, F.; Saygin, D.; Bazilian, M.D.; Wagner, N.; Gorini, R. The role of renewable energy in the global energy transformation. *Energy Strategy Rev.* **2019**, *24*, 38–50. [\[CrossRef\]](#)
3. Avancini, B.; Rodrigues, C.; Martins, B.; Rabêlo, L.; Al-Muhtadi, J.; Solic, P. Energy meters evolution in smart grids: A review. *J. Clean. Prod.* **2019**, *217*, 702–715. [\[CrossRef\]](#)
4. Quilumba, L.; Lee, W.; Huang, H.; Wang, Y.; Szabados, L. Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. *IEEE Trans. Smart Grid* **2014**, *6*, 911–918. [\[CrossRef\]](#)
5. Kipping, A.; Trømborg, E. Modeling aggregate hourly energy consumption in a regional building stock. *Energies* **2018**, *11*, 78. [\[CrossRef\]](#)
6. Shi, H.; Xu, M.; Li, R. Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Trans. Smart Grid* **2017**, *9*, 5271–5280. [\[CrossRef\]](#)
7. Edwards, E.; New, J.; Parker, E. Predicting future hourly residential electrical consumption: A machine learning case study. *Energy Build.* **2012**, *49*, 591–603. [\[CrossRef\]](#)
8. Arvanitidis, A.I.; Bargiotas, D.; Daskalopulu, A.; Kontogiannis, D.; Panapakidis, I.P.; Tsoukalas, L.H. Clustering Informed MLP Models for Fast and Accurate Short-Term Load Forecasting. *Energies* **2022**, *15*, 1295. [\[CrossRef\]](#)
9. Ghofrani, M.; Hassanzadeh, M.; Etezadi-Amoli, M.; Fadali, S. Smart meter based short-term load forecasting for residential customers. In Proceedings of the North American Power Symposium, Boston, MA, EUA, 4–6 August 2011. [\[CrossRef\]](#)
10. Park, S.; Jung, S.; Jung, S.; Rho, S.; Hwang, E. Sliding window-based LightGBM model for electric load forecasting using anomaly repair. *J. Supercomput.* **2021**, *77*, 27–30. [\[CrossRef\]](#)
11. Kong, W.; Dong, Z.Y.; Hill, D.J.; Luo, F.; Xu, Y. Short-term residential load forecasting based on resident behaviour learning. *IEEE Trans. Power Syst.* **2018**, *33*, 1087–1088. [\[CrossRef\]](#)
12. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* **2019**, *10*, 841–851. [\[CrossRef\]](#)
13. Alves, M.F. Previsão de Cargas Não Residenciais Mistas Por Redes Neurais Artmap Fuzzy. Ph.D. Thesis, Universidade Estadual Paulista, Ilha Solteira, Brazil, 2019.
14. Haq, E.U.; Lyu, X.; Jia, Y.; Hua, M.; Ahmad, F. Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach. *Energy Rep.* **2020**, *6*, 1099–1105. [\[CrossRef\]](#)
15. Moon, J.; Jung, S.; Rew, J.; Rho, S.; Hwang, E. Combination of short-term load forecasting models based on a stacking ensemble approach. *Energy Build* **2020**, *216*, 109921. [\[CrossRef\]](#)
16. Ayub, N.; Irfan, M.; Awais, M.; Ali, U.; Ali, T.; Hamdi, M.; Alghamdi, A.; Mahammad, F. Big Data Analytics for Short and Medium-Term Electricity Load Forecasting Using an AI Techniques Ensembler. *Energies* **2020**, *13*, 5193. [\[CrossRef\]](#)
17. Roth, J.; Chadalawada, J.; Jain, R.K.; Miller, C. Uncertainty Matters: Bayesian Probabilistic Forecasting for Residential Smart Meter Prediction, Segmentation, and Behavioral Measurement and Verification. *Energies* **2021**, *14*, 1481. [\[CrossRef\]](#)
18. Jung, S.; Moon, J.; Park, S.; Hwang, E. An Attention-Based Multilayer GRU Model for Multistep-Ahead Short-Term Load Forecasting. *Sensors* **2021**, *21*, 1639. [\[CrossRef\]](#)
19. Tkachenko, R. An Integral Software Solution of the SGTm Neural-like Structures Implementation for Solving Different Data Mining Tasks. *Lect. Notes Data Eng. Commun. Technol.* **2022**, *77*, 696–713. [\[CrossRef\]](#)
20. Izonin, I.; Tkachenko, R.; Kryvinska, N.; Tkachenko, P.; Greguš ml, M. Multiple Linear Regression Based on Coefficients Identification Using Non-iterative SGTm Neural-like Structure. *Lect. Notes Comput. Sci.* **2019**, *11506*, 467–479. [\[CrossRef\]](#)
21. Prakash, K.; Sydulu, M. Non iterative-state estimation based neural network for short term load forecasting of distribution systems. In Proceedings of the 2009 IEEE Power & Energy Society General Meeting, Calgary, AB, Canada, 26–30 July 2009; pp. 1–8. [\[CrossRef\]](#)
22. Ishwaran, H.; Rao, J.S. Spike and slab variable selection: Frequentist and bayesian strategies. *Ann. Stat.* **2005**, *33*, 730–773. [\[CrossRef\]](#)
23. Tkachenko, R.; Izonin, I. Model and Principles for the Implementation of Neural-like Structures Based on Geometric Data Transformations. *Adv. Intell. Syst. Comput.* **2019**, *754*, 578–587. [\[CrossRef\]](#)
24. Cover, M.; Thomas, A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006.
25. Kim, S.H.; Geem, Z.W.; Han, G.T. Hyperparameter Optimization Method Based on Harmony Search Algorithm to Improve Performance of 1D CNN Human Respiration Pattern Recognition System. *Sensors* **2020**, *20*, 3697. [\[CrossRef\]](#)
26. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
27. Jain, R.K.; Smith, K.M.; Culligan, P.J.; Taylor, J.E. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Appl. Energy* **2014**, *123*, 168–178. [\[CrossRef\]](#)
28. Gajowniczek, K.; Zabkowski, T. Short term electricity forecasting using individual smart meter data. *Procedia Comput. Sci.* **2014**, *35*, 589–597. [\[CrossRef\]](#)