Journal of Scientific & Industrial Research
Vol. 83, January 2024, pp. 58-66
DOI: 10.56042/jsir.v83i1.3523

# Monthly Electricity Consumption Prediction: Integrating Artificial Neural Networks and Calculated Attributes

Knežević Dragana[1]*, Blagojević Marija[2] & Ranković Aleksandar[2]
[1]Western Serbia Academy of Applied Studies, Trg Svetog Save 34, Užice 31000, Serbia
[2]University of Kragujevac, Faculty of Technical Sciences Čačak, Svetog Save 65, 32102 Čačak, Serbia

*Received 01 July 2023; revised 26 December 2023; accepted 28 December 2023*

Electricity consumption is increasing on a daily basis, and consequently, the need for its control, potential reducing or at least predicting, is growing. The aim of this research is to predict the electricity consumption based on consumer attributes, using a dataset with a poor list of useful attributes as a starting point. Even though the electricity distribution company from which the data were obtained records data on electricity consumption precisely, the obtained data did not provide enough information to ensure a satisfactory level of the estimation precision. That is why, for the purpose of this research, the initial dataset was subjected to the extensive treatment in the preprocessing phase and updated with a lot of additional, collected and calculated attributes. Subsequently, the neural network model that predicts electricity consumption on a monthly basis was proposed. Basically, two models were created, with several variations in the number of neurons in the hidden layers, but with the identical structure of input and output layers. The proposed models were tested on a very complex dataset, obtained by updating the initial one, and comprising all the measuring points and all types of consumers in the area of the City of Užice, recorded during a period of 56 months. The results show that the proposed metodology of updating a dataset with additionaly collected and calculated inputs, together with the proposed neural network model, ensures a very low prediction error, i.e., ≈5%. This could make electricity consumption control and reduction, but also electricity production planning possible.

**Keywords:** ANN, Consumer, Data mining, Layer normalization layers, Weight normalization layers

## Introduction

The prediction of electricity consumption is becoming inevitable for many reasons and significant for both electricity consumers and producers. Electrical energy is inconvenient to store and therefore it is very important to predict electricity consumption as precisely as possible. This is why the number of authors working on the solution to this problem using different machine learning techniques is increasing. The authors of the existing papers mostly deal with short-term electricity consumption predicting, based on the data recorded over a short period of time within a specific target area.[1] On the other hand, this research aims to predict the electricity consumption on a montly basis based on a dataset containing the information on the consumption by all the consumers within the territory of an entire city.

Due to the complex nature of the dataset, it was quite challenging to achieve a satisfactory level of electricity consumption prediction accuracy.

*Author for Correspondence
E-mail: dragana.knezevic@vpts.edu.rs; dknezevic28@gmail.com

This research has two main goals: to create a neural network model capable to estimate the monthly electricity consumption based on real, recorded data, and to make the initial dataset suitable for the exploitation by updating it.

The research will serve as a starting point for further research of the kind, as it provides the information that will ensure successful and even more precise use of the conventional techniques and algorithms.

## Literature Review

So far, there has been extensive research on the use of data mining techniques for electricity consumption classification and predicting. Some authors carried out comparative reviews and analyses of different data mining techniques used to predict electricity consumption[2–4], while others approached the problem from the economic perspective.[5,6]

Electricity consumption predicting has been performed in different environments and time frames. Some authors used neural networks for long-term electricity consumption predicting[7–12], but they are by far outnumbered by those who used them for short-time predicting, ranging between a few minutes and a few hours or days.[1,4,12–17]

Many authors used neural networks to make predictions for a specific area, country[8], city[18,19], or even a shopping mall[9], power substation or building.[20]

For the purpose of this paper, the power consumption information was obtained by manually recording the amount of kWh (Kilowatt-hour) from the monthly readings of the electric energy meters.[18,21]

**Materials and Methods**

The above-given review of the existing literature in this field indicates the original nature of the research carried out for the purpose of this paper. Namely, this research aims to address the problem by preparing the data and developing a predictive model for a specific territory with different types of users and, at the same time, creating a model convenient for processing other sets with similar characteristics.

Despite being recorded by the electricity distribution company, the electricity consumption data did not provide enough information to ensure a satisfactory level of the estimation precision. Therefore, it was quite challenging to create the final, valid set of input data, which would ensure much more precise prediction of electricity power consumption.

The main methodology of this research involved identifying the appropriate real-world attributes, collected from different sources and calculated based on the other available attributes, and creating a neural network model capable to predict the electricity consumption.

The form of the real-world, raw data was not well suited for further exploitation, and as such, had to be thoroughly processed in the preprocessing phase: irrelevant attributes were excluded from the dataset (the readings from electricity meters for each tariff, at the beginning and end of a month, and reading dates). Ultimately, only a few attributes remained in the set (hereinafter: ‘the initial set’), such as: *Consumer label, Consumer Category, Zone, Consumer Group, Calculation Period* and *Consumption in kWh*. Such an initial set was very poor in useful information and therefore was updated with the data subsequently collected from various sources (the electricity distribution company and weather station) for each month within the observed period:

- The meteorological data including *Average temperature (℃), Humidity (%), Precipitation (mm), Number of days with precipitation, Max temperature(℃), Min temperature(℃), Number of clear days, Number of cloudy days, Insolation in hours, Sunshine hours* and *Daylight hours* were collected from Republic Hydrometeorological Service of Serbia[22];
- The number of *Work days, Weekends* and *Holidays,* was collected for every single month.

Besides the above mentioned list of collected data, the dataset was updated with additional, calculated attributes, for each month within the observed period:

- *Summer/Winter Time* – describes the time of the year when the country observes Daylight Saving Time (DST), winter time - from November to March or summer time - from April to October;
- *Seasons* – four divisions of the year according to annual weather changes (winter, spring, summer and autumn), based on the calculation period;
- Attributes *1month Before, 2months Before, 3months Before* were calculated as lagged dependent variables (LDVs) for each consumer[23], and represent the consumption that each consumer had had in the previous three months;
- Attributes *Consumption Zone* and *Consumption Zone 1 MB* were determined based on the number of consumed kWh in the previous month or the month before, respectively. The consumption between 350 and 1600 kWh belongs to the blue zone, the consumption below 350 kWh to the green zone, and the consumption of over 1600 kWh to the red zone;
- *Average consumption* represents the average consumption per consumer, taking into account only the consumption within the observed period;
- The attribute *Trend* was calculated using a linear regression function, the least squares method (Eq. (1)), and by processing the consumption by each consumer separately as a time series:

$$Y_i = \beta_0 + \beta_1 X_i \qquad \dots (1)$$

where, $Y_i$ is the dependent variable – the consumption in kWh, $X_i$ is the independent, explanatory variable (the time index for each consumer), $\beta_0$ is the intercept (the value of $Y$ when $X = 0$), and $\beta_1$ is the slope of the line;

- The attribute *Scores* was calculated using the Isolation Forest method[24,25], which was also used to discover anomalies within the dataset itself, and represents the anomaly score of an instance. The anomaly score is defined as Eq. (2). Its values range between 0 and 1, and the values closer to 0 indicate an anomaly:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \quad \dots (2)$$

where, $h(x)$ is the path length of observation $x$, $E(h(x))$ is the average of $h(x)$, $n$ is the number of external nodes, $c(n)$ is the average of $h(x)$ given $n$, ($c(n)$ is used to normalize $h(x)$).

- Attributes *Difference from max consumption* and *Difference from min consumption* were calculated as the difference between each recorded consumption in kWh and the maximum and minimum consumption per consumer in the observed period.

After defining the inputs and before creating an Multi Layer Perceptron (MLP) model, categorical variable encoding and data scaling were performed. There were two scenarios, for which two different types of categorical variable encoding were used: in the first scenario, Mean Encoding (Target Encoding)[26] was used to encode all categorical variables. In the second scenario, One-Hot Encoding was used for all the categorical variables except for the *Consumer label*, for which, due to numerous possible categories, Target encoding was used. The data encoding and scaling were followed by the selection of the model type. There were two potential model types:

- *model type 1:* two hidden dense Weight Normalization layers (marked as WNL in Fig. 1)[27] and two Layer Normalization Layers (marked as LNL in Fig. 1) (Jimmy *Lei Ba et al.*, Layer normalization, unpublished data, published online 2016) (after the first hidden layer and before the output layer) or
- *model type 2:* two basic dense layers.

Both models used scaling variance with uniform distribution for initial random weights, He Normal[28] for bias initialization, Elastic Net regularization[29] to avoid overfitting, and Adadelta optimizer with a reducing learning rate over training epochs. The mean squared error was used as a loss function. Both models were tested on 30% dataset, using the same evaluation measures: the Mean of the Absolute Percentage Errors of prediction (MAPE), Root Mean Squared Error (RMSE), relative RMSE, and the coefficient of determination (Rsquared). The proposed methodology comprised several steps, which are graphically represented by the flow chart in Fig. 1.

**Multi Layer Perceptron (MLP)**

MLP is the most commonly used feedforward neural network, primarily because of its fast operation and ease of implementation.[30]

MLPs are the quintessential deep learning models, which use a supervised learning algorithm that learns a function $f()$: $R_n \rightarrow R_o$ by training on a dataset in the offline batch regime, where $n$ is the number of

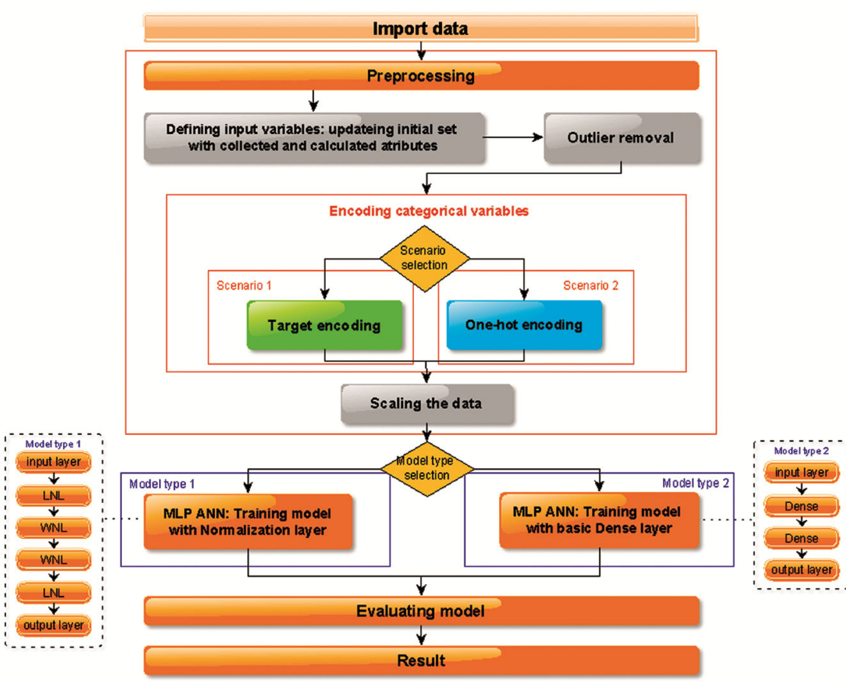


Fig. 1 — Flow chart of the proposed methodology

dimensions of the input and *o* is the number of dimensions for the output.

A feedforward network defines a mapping $Y = f(X;W)$ and learns the value of the parameters represented by $W$ matrix, which results in the best function approximation.[31]

An MLP is a finite directed acyclic graph where, for each neuron *j* in the hidden layer, we sum its input signals $x_i$ impinging onto it after multiplying them by their respective connection weights $w_{ji}$. The output of each neuron is described as follows[32]:

$$y_j = f(\sum w_{ji} x_i) \quad \ldots (3)$$

where, *f* is an activation function applied to the weighted sums of the inputs.

The model used in this paper has a structure where each neuron in one layer is connected to all neurons in the next layer without shortcuts. All connections are weighted with a real number. The weight of the connection $i \rightarrow j$ is called $w_{ij}$. All hidden and output layers have a weight bias. For *i* neuron, weight bias is called $w_{i0}$.

The weight change in every hidden layer, before the output one, is calculated using the following equation:

$$w_{ji}(k+1) = w_{ji}(k) + \mu f'\left(net_j(k)\right)\left(\sum_a \varepsilon_a(k) f'\left(net_a(k)\right) w_{aj}(k)\right) u_{ji} \quad \ldots (4)$$

here, $\left(\sum_a \varepsilon_a(k) f'\left(net_a(k)\right) w_{aj}(k)\right)$ is the sum of all local errors of all output neurons, multiplied by the weights. In this paper, the nodes in the input layer used a standard sigmoid activation function (Fig. 2(a), Eq. (5)), those in the hidden layers used a ReLu activation function (Fig. 2(b), Eq. (6)), whereas for the output layer, the Linear activation function was used (Fig. 2(c), Eq. (7)).

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad \ldots (5)$$

$$R(x) = \max(0, x) \quad \ldots (6)$$

$$f(x) = x \quad \ldots (7)$$

In Fig. 3 the MLP structure used in this study is shown, which comprises three main parts: the input layer, hidden layers and output layer. The input layer consists of all the attributes kept in the initial set with the collected and calculated attributes added, i.e., a total of 31 attributes. The hidden layers in this case have dozens of neurons and the output layer has one neuron (the target attribute – monthly electricity consumption in kWh). In this research, two different model types, mentioned above, were used.

## Results and Discussion

Once it was updated, the target dataset included data about electricity consumers on the territory of the City of Užice for the period of four years and eight months. The set comprised over 1.5 milion records for a total of around 40,000 consumers, thus covering all the existing measurement points on this specific territory. The consumers who live on this territory range between those with low monthly consumption (such as small households) to those whose consumption reaches thousands of kWh (such as large non-households), which implies that the target dataset has a very heterogeneous structure. Different divisions of consumers included in the dataset are shown, according to different criteria: the zone in which consumers live (a), the category of consumers (b) and the group of consumers (c), are shown in Fig. 4.

The data were processed and the neural network modelled in Python, a programming language. To evaluate the model, a few measures were used. Root Mean Square Error (RMSE) is widely used for the evaluation of the performance of a model. [11,33] RMSE

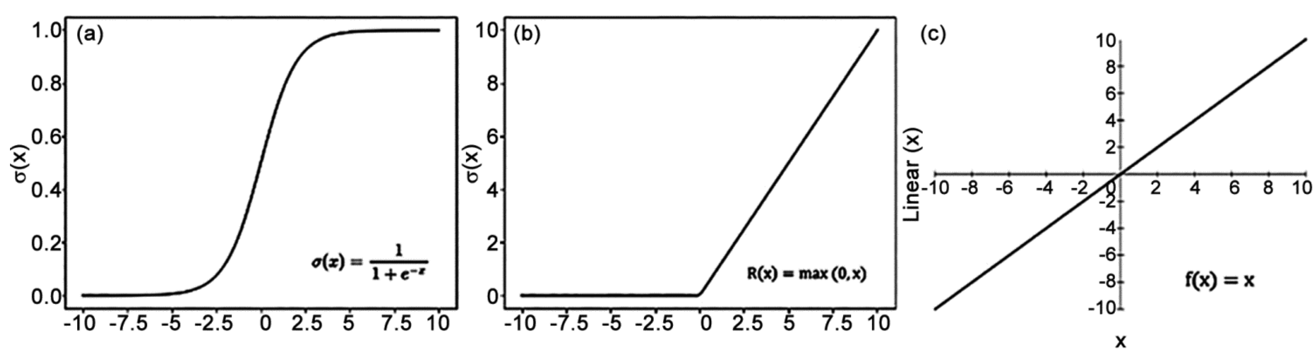


Fig. 2 — Activation function used for input layer: (a) sigmoid, (b) for hidden layers - ReLu and (c)for output layer - Linear

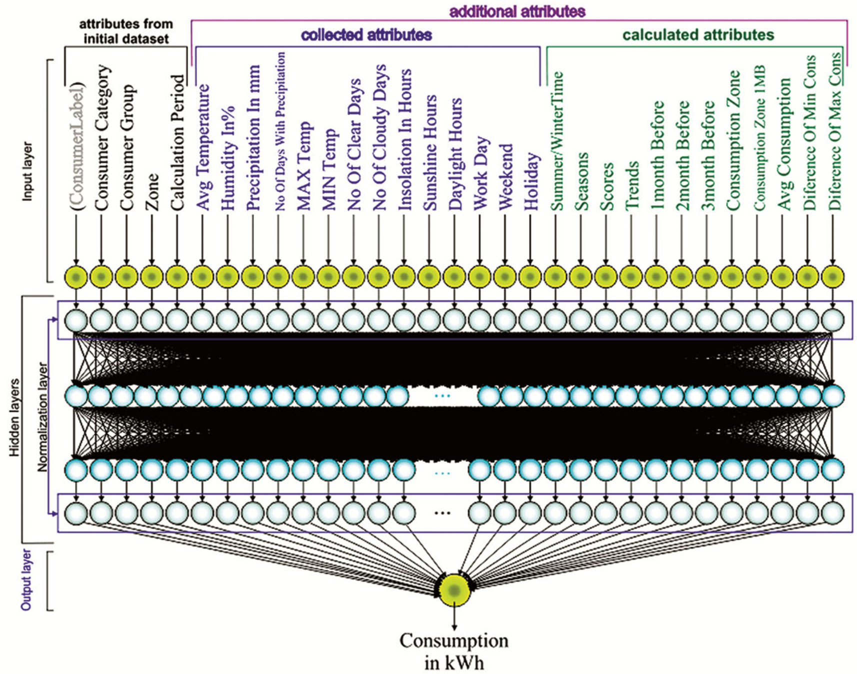


Fig. 3 — The structure of the MLP model(s) used for monthly electricity consumption prediction

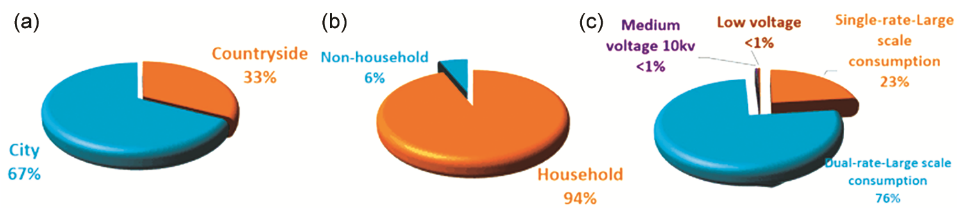


Fig. 4 — Overview of consumers structure in the City of Užice by different criteria

measures the standard deviation of residuals (Eq. (8)):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{actual}_i - \mathrm{prediction}_i)^2} = \sqrt{\mathrm{MSE}} \quad \dots (8)$$

In the previous equation $n$ is the number of samples, $e_i$ is the prediction error, $prediction_i$ is the value of the output parameter $i$ of the network, and $actual_i$ is the accurate value of the output parameter $i$.

Based on the absolute RMSE value, we can calculate the relative RMSE (RRMSE or RMSE (%)) Eq. (9):

$$\mathrm{RMSE}\ (\%) = \frac{\mathrm{RMSE}}{\text{mean count}} * 100 \quad \dots (9)$$

MAPE (Mean Absolute Percentage Error) is another measure used in this paper for the evaluation of the model. MAPE represents the mean or average of the absolute percentage errors of prediction:

$$\mathrm{MAPE}\ (\%) = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\mathrm{actual}_i - \mathrm{prediction}_i}{\mathrm{actual}_i}\right| * 100 \quad \dots (10)$$

Finally, one more measure used, the coefficient of determination, or $R^2$ (Rsquared), measures how well the regression line approximates the actual data (Eq. (11)).

$$R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}} = 1 - \frac{\sum(\mathrm{actual}_i - \mathrm{prediction}_i)^2}{\sum(\mathrm{actual}_i - \mathrm{mean})^2} \quad \dots (11)$$

The results of the proposed models, evaluated using these measures, are described in this chapter, including both scenarios and model types mentioned above, and together with the difference in the result when the same model is used with and without the consumer label.

The initial set, which consisted of only five useful attributes, did not provide enough information for satisfactory prediction results. The average MAPE error for the initial set was ≈70%, and the average

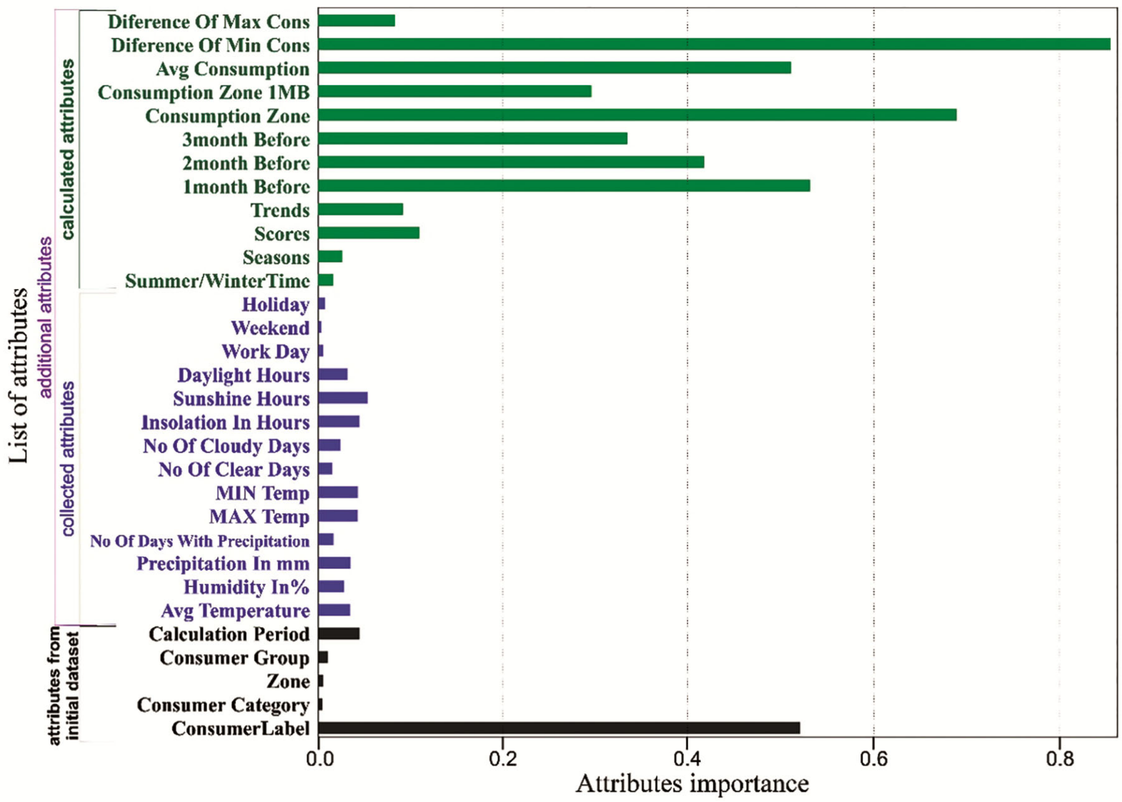


Fig. 5 — Importance of attributes in final dataset

value of RMSE was ≈147 kWh. It is clear that such results had to be improved by including additional attributes.

The importance of additional attributes in relation to the target variable (Consumption in kWh) compared to the attributes from the initial set, is shown in Fig. 5. All the attributes from the initial set have very low significance, except for the *Consumer label*, which plays an important role due to the nature of the data set. This indicates that we should use this attribute as an input variable, but due to the large number of existing consumer labels, this can be done only by using Target encoding.

In Table 1 the results given by various Artificial Neural Network (ANN) models, created for the new, updated dataset are shown. In order to avoid overfitting, the Elastic Net regularization and different parameters were used:

- different number of neurons in hidden layers,
- different scenarios (see Fig. 1),
- different model types (see Fig. 1) and
- different inputs, i.e., with or without the *Consumer label* attribute.

The results given in Table 1 show very high precision of all proposed scenarios and ANN model types, as well as that the number of neurons in the hidden layers plays a very important role. In general, better results, as expected, are obtained using the model with more neurons (65 or 85 neurons) in hidden layers. The best result was obtained by the *scenario 2* + *model type 1,* when the *Consumer label* was used as an input.

The number of epochs in ANN training is a very important factor. More epochs require more computer resources and more execution time. The number of training epochs by models varies between 14 and 150, as shown in Fig. 6. The number of epochs was lower when Target encoding was applied, but not directly connected with the number of neurons in the hidden layers (Fig. 6 – for all variants of *scenario 1*, marked in blue: the model with Target encoding; for all variants of *scenario 2*, marked in orange: the model with One-hot encoding). Although it provided better results, the model in which One-hot encoding was applied (*scenario 2*) needed more training epochs than the model in which Target encoding was applied (*scenario 1*). Generally, the model which combined the *scenario 2* and *model type 2*, i.e., the model to which One-hot encoding was applied and the model with basic dense hidden layers, required the greatest number of epochs.

Table 1 — Performance measure obtained by different scenarios and ANN model types

| ANN layers | Evaluation measures | *Scenario* 1 + *model type* 1 With consumer label | *Scenario* 1 + *model type* 1 Without consumer label | *Scenario* 1 + *model type* 2 With consumer label | *Scenario* 1 + *model type* 2 Without consumer label | *Scenario* 2 + *model type* 1 With consumer label | *Scenario* 2 + *model type* 1 Without consumer label | *Scenario* 2 + *model type* 2 With consumer label | *Scenario* 2 + *model type* 2 Without consumer label |
|---|---|---|---|---|---|---|---|---|---|
| With 85 neurons in each layer | RMSE (kWh) | 29.35 | 30.26 | 30,97 | 32.64 | 25.1 | 30.008 | 26.08 | 30.93 |
| | RMSE (%) | 7.02 | 7.24 | 7.4 | 7.81 | 6.00 | 7.18 | 6.19 | 7.40 |
| | MAPE (%) | 5.15 | 5.34 | 5.40 | 5.59 | 4.11 | 5.28 | 4.19 | 5.35 |
| | Rsquared | 0.979 | 0.978 | 0.976 | 0.974 | 0.985 | 0.979 | 0.983 | 0.977 |
| With 65 neurons in each layer | RMSE (kWh) | 31.22 | 30.39 | 30.99 | 33.18 | 25.14 | 30.06 | 26.4 | 30.73 |
| | RMSE (%) | 7.47 | 7.27 | 7.41 | 7.94 | 6.01 | 7.19 | 6.31 | 7.35 |
| | MAPE (%) | 5.49 | 5.38 | 5.47 | 5.95 | 4.12 | 5.32 | 4.38 | 5.48 |
| | Rsquared | 0.976 | 0.977 | 0.977 | 0.973 | 0.985 | 0.978 | 0.983 | 0.977 |
| With 35 neurons in each layer | RMSE (kWh) | 30.76 | 31.21 | 31.47 | 32.2 | 25.78 | 30.13 | 26.775 | 31.44 |
| | RMSE (%) | 7.36 | 7.46 | 7.53 | 7.70 | 6.17 | 7.21 | 6.40 | 7.52 |
| | MAPE (%) | 5.38 | 5.54 | 5.54 | 5.74 | 4.25 | 5.34 | 4.42 | 5.67 |
| | Rsquared | 0.977 | 0.976 | 0.976 | 0.975 | 0.984 | 0.978 | 0.983 | 0.976 |
| With 15 neurons in each layer | RMSE (kWh) | 32.67 | 33.79 | 34.19 | 34.89 | 27.55 | 31.68 | 27.86 | 34.14 |
| | RMSE (%) | 7.81 | 8.08 | 8.18 | 8.34 | 6.66 | 7.58 | 6.66 | 8.26 |
| | MAPE (%) | 5.7 | 6.02 | 6.16 | 6.31 | 4.588 | 5.7 | 4.62 | 6.17 |
| | Rsquared | 0.974 | 0.982 | 0.971 | 0.97 | 0.981 | 0.975 | 0.981 | 0.972 |

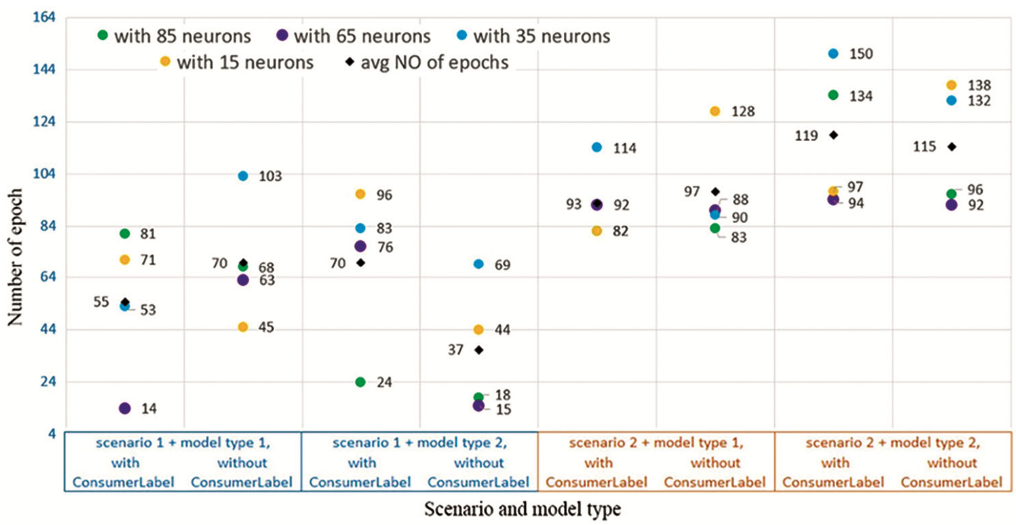


Fig. 6 — Number of epochs by different scenarios/model/label types and different number of neurons

To get a clearer picture of the findings, Fig. 7 shows the distribution of MAPE values in 4 different cases (with the Consumer label) for the models with 85 neurons in the hidden layer yielding the best results according to Table 1. In each case, some consumers have extremely high MAPE values. The blue and red line represent the limits of the range of ±15% of MAPE. The consumers outside the range make up only a small percentage of the total number of consumers in the tested dataset. The number of such consumers is the smallest in the scenario 2 + model type 1 (2.28%), whereas it is the highest in the scenario 1 + model type 2 (4.87%). Finally, with 97% of consumers, the error is lower than 15%.

The results obtained using the proposed models show that electricity consumption can be predicted on a monthly level in a heterogeneous dataset which includes all the consumers within the area of a city with a small prediction error (≈25 kWh). The minimal MAPE value of ≈4.1% represents a much better result than those obtained in the previous, similar research.

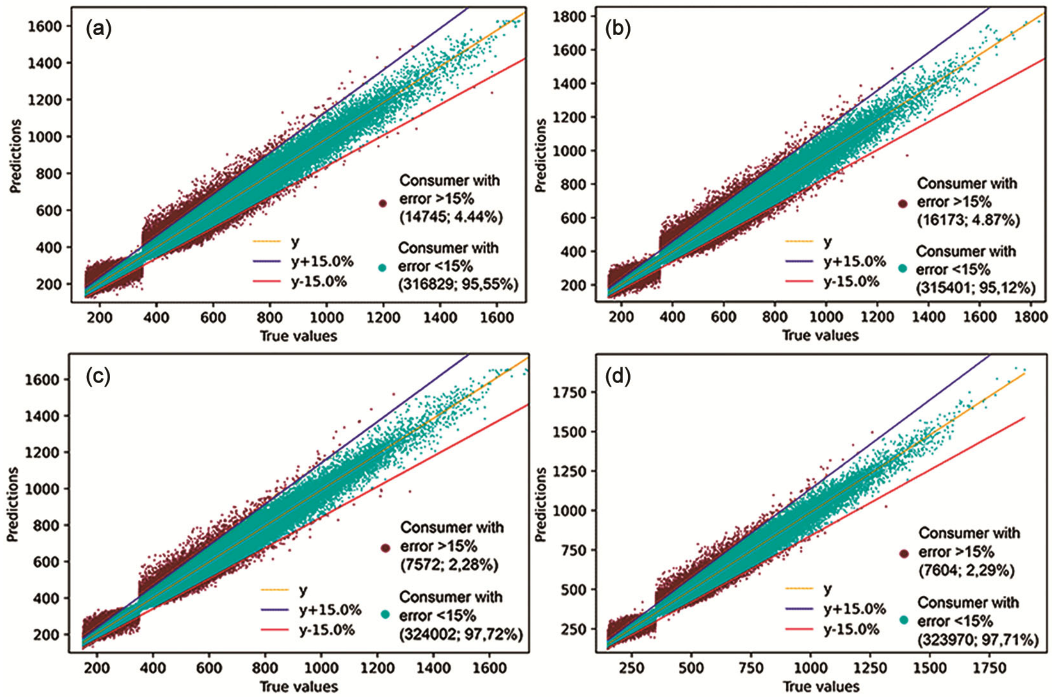


Fig. 7 — Comparison of actual values and those predicted using different ANN scenarios and model types (with Consumer label) within ±15% error range: (a) Scenario 1 + model type 1, with consumer level; (b) Scenario 1 + model type 2, with consumer level; (c) Scenario 2 + model type 1, with consumer level; (d) Scenario 2 + model type 2, with consumer level

## Conclusions

Finding the way to predict electricity consumption has become one of the most important tasks of the research in this area. Therefore, the results of this study are important for both producers and consumers because they guarantee the reliability of electricity consumption predicting in real-world conditions. They can also help to plan the production and distribution of electricity at the level of a single distribution center, and to predict the consumption by new consumers in an area with similar characteristics.

The obtained results and findings can serve as the stepping stone for further research. Using data similar to those presented in the paper, future studies can draw upon these techniques and algorithms to predict the consumption on a monthly basis in any city, country or region with high reliability.

Future research could even further improve the results by starting from a homogenous dataset, i.e., by taking only specific groups or categories of consumers into consideration, but also improve the models by adding new layers of neurons and adjusting other parameters.

## Acknowledgement

This study was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, and these results are part of the Grant No. 451-03-68/2022-14/200132 with the University of Kragujevac – Faculty of Technical Sciences Čačak.

## References

1 Chae Y T, Horesh R, Hwang Y & Lee Y M, Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings, *Energy Build*, **111** (2016) 184–194, doi: http://dx.doi.org/10.1016/j.enbuild.2015.11.045.

2 Tso G K F & Yau K K W, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks, *Energy*, **32**(**9**) (2007) 1761–1768, doi: https://doi.org/10.1016/j.energy.2006.11.010.

3 Aman S, Frincu M, Charalampos C, Noor U, Simmhan Y & Prasanna V, Empirical comparison of prediction methods for electricity consumption forecasting, *USC*, Published online 2014 14-942.

4 Hippert H S, Pedreira C E & Souza R C, Neural networks for short-term load forecasting: A review and evaluation, *IEEE Trans Power Syst*, **16**(**1**) (2001) 44–55, doi: 10.1109/59.910780.

5 Hu Y C, Electricity consumption prediction using a neural-network-based grey forecasting approach, *J Oper Res Soc*, **68**(**10**) (2017) 1259–1264, doi: https://doi.org/10.1057/s41274-016-0150-y.

6 Song X, Liang G, Li C & Chen W, Electricity consumption prediction for Xinjiang electric energy replacement, *Math Probl Eng*, **2019** (2019), doi: https://doi.org/10.1155/2019/3262591.

7 Yoo S G & Myriam H Á, Predicting residential electricity consumption using neural networks: A case study, *J Phys Conf Ser*, **1072** (2018), doi: 10.1088/1742-6596/1072/1/012005.

8 Kargar M J & Charsoghi D, Predicting annual electricity consumption in Iran using artificial neural networks (NARX), *Indian J Sci Res*, **5**(**1**) (2014) 231–242, doi: 10.1080/15567240802706734.

9 Kim M K, Cha J, Lee E, Pham V H, Lee S & Theera-Umpon N, Simplified neural network model design with sensitivity analysis and electricity consumption prediction in a commercial building, *Energies*, **12(7)** (2019) 1201, doi: 10.3390/en12071201.

10 Kavaklioglu K, Ceylan H, Ozturk H K & Canyurt O E, Modeling and prediction of Turkey's electricity consumption using artificial neural networks, *Energy Convers Manag*, **50**(**11**) (2009) 2719–2727, doi: https://doi.org/10.1016/j.enconman.2009.06.016.

11 Azadeh A, Ghaderi S F, Tarverdian S & Saberi M, Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption, *Appl Math Comput*, **186**(**2**) (2007) 1731–1741, doi: 10.1109/IECON.2006.348098.

12 Berriel R F, Lopes A T, Rodrigues A, Varejao F M & Oliveira-Santos T, Monthly energy consumption forecast: A deep learning approach, *Proc Int Jt Conf Neural Netw*, (**May**) (2017) 4283–4290, doi: 10.1109/IJCNN.2017.7966398.

13 Azadeh A, Saberi M, Nadimi V, Iman M & Behrooznia A, An integrated intelligent neuro-fuzzy algorithm for long-term electricity consumption: cases of selected EU countries, *Acta Polytech Hung*, **7(4)** (2010) 71–90, doi: 10.1109/EMS.2010.56.

14 Selakov A, Cvijetinović D & Ilić S, Short term load forecasting using support vector machines for different consumer types, *J Process Energy Agric*, **16**(**3**) (2012) 128–131.

15 Ilić S, Vukmirović S, Erdeljan A & Kulić F, Hybrid artificial neural network system for short-term load forecasting, *Therm Sci*, **16**(**suppl. 1**) (2012) 215–224, doi: 10.2298/TSCI120130073I.

16 Amjady N & Keynia F, Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm, *Energy*, **34**(**1**) (2009) 46–57, doi: https://doi.org/10.1016/j.energy.2008.09.020.

17 Banga A, Ahuja R & Sharma S, Stacking machine learning models to forecast hourly and daily electricity consumption of household using internet of things, *J Sci Ind Res*, **80** (2021) 894–904.

18 Knežević D & Blagojević M, Classification of electricity consumers using artificial neural networks, *FU Elect Energ*, **32**(**4**) (2019) 529–538, doi: 10.2298/FUEE1904529K.

19 Beccali M, Cellura M, Brano V L & Marvuglia A, Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area, *Renew Sust Energ Rev*, **12**(**8**) (2008) 2040–2065, doi: https://doi.org/10.1016/j.rser.2007.04.010.

20 Cai H, Shen S, Lin Q, Li X & Xiao H, Predicting the energy consumption of residential buildings for regional electricity supply-side and demand-side management, *IEEE Access*, **7** (2019) 30386–30397, doi: 10.1109/ACCESS.2019.2901257.

21 Rathod R R & Garg R D, Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data, *Int J Electr Power Energy Syst*, **78** (2016) 368–374, doi: https://doi.org/10.1016/j.ijepes.2015.11.110.

22 RHMZ - Republic Hydrometeorological Service of Serbia, Accessed January 18, 2022. http://www.hidmet.gov.rs/latin/meteorologija/klimatologija_godisnjaci.php.

23 Wilkins A S, To lag or not to lag?: Re-evaluating the use of lagged dependent variables in regression analysis, *Political Sci Res Methods*, **6**(**2**) (2018) 393–411, doi: https://doi.org/10.1017/psrm.2017.4.

24 Liu F T, Ting K M & Zhou Z H, Isolation-based anomaly detection, *ACM Transact Knowl Discov Data (TKDD)*, **6**(**1**) (2012) 1–39, doi: https://doi.org/10.1145/2133360.2133363.

25 Ding Z & Fei M, An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window, *IFAC Proc*, **46**(**20**) (2013) 12–17, doi: https://doi.org/10.3182/20130902-3-CN-3020.00044.

26 Pargent F, Pfisterer F, Thomas J & Bischl B, Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features, *Comput Stat*, **37** (2022) 2671–2692, doi: https://doi.org/10.1007/s00180-022-01207-6.

27 Salimans T & Kingma D P, Weight normalization: A simple reparameterization to accelerate training of deep neural networks, *Adv Neural Inf Process*, **29** (2016) 901–909, doi: https://doi.org/10.48550/arXiv.1602.07868.

28 He K, Zhang X, Ren S & Sun J, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, *Proc IEEE Int Conf*, (2015) 1026–1034, doi: 10.48550/arXiv.1502.01852.

29 Zou H & Hastie T, Regularization and variable selection via the elastic net, *J R Stat Soc B: Stat Methodol*, **67** (**2**) (2005) 301–320, doi: https://doi.org/10.1111/j.1467-9868.2005.00503.x.

30 Haykin S S, *Neural Networks and Learning Machines/Simon Haykin*, 3rd ed, (New York: Prentice Hall), 2009.

31 Aleksić S, Pantić A & Pantić D, High electric field stress model of n-channel VDMOSFET based on artificial neural network, *J Comput Electron*, **17**(**3**) (2018) 1210–1219, doi: 10.1007/s10825-018-1167-z.

32 Orhan U, Hekim M & Ozer M, EEG signals classification using the K-means clustering and a multilayer perceptron neural network model, *Expert Syst Appl*, **38**(**10**) (2011) 13475–13481, doi: https://doi.org/10.1016/j.eswa.2011.04.149.