Electricity Consumption Prediction Model for Improving Energy Efficiency Based on Artificial Neural Networks

Dragana KNEŽEVIĆ[1]*, Marija BLAGOJEVIĆ[2], Aleksandar RANKOVIĆ[2]

[1] Western Serbia Academy of Applied Studies, Užice 31000, Trg Svetog Save 34, Serbia
dknezevic28@gmail.com (**Corresponding author*)

[2] University of Kragujevac, Faculty of Technical Sciences Čačak, 32102 Čačak, Svetog Save 65, Serbia
marija.blagojevic@ftn.kg.ac.rs, aleksandar.rankovic@ftn.kg.ac.rs

Abstract: Continuous population growth is causing an increasing electricity demand. In order to provide enough electricity, it should be possible to predict the prospective consumption. This is especially important nowadays, when energy-saving measures aimed at improving the energy efficiency of all energy sources, especially electrical ones, are gaining importance. Neural networks play an important role in predicting electricity consumption. This paper aims to provide the neural network architecture that will facilitate the prediction of the monthly consumption of different types of consumers with a minimum error. The proposed model is based on two uncommon types of layers, and its reliability is tested on a real dataset related to the electricity consumption of all consumers on the territory of the City of Užice in Serbia. To ensure that more precise results are obtained, this paper also sets forth another approach involving the dataset partitioning into meaningful units (subclusters) before applying the proposed model to them. Finally, the architecture of the Electricity Consumption Prediction System (ECPS) is presented, as an interactive GUI intended for the end user. The dataset employed for training the implemented models contains the consumption data collected over a period of three years, whereas the test set contains data from the fourth year, which corresponds to the actual conditions in which the application will be used.

Keywords: Neural network, Electricity consumption, Prediction, Custom neural network model, Electricity Consumption Prediction System (ECPS).

1. Introduction

The prediction of electricity consumption has been gaining importance over the last couple of years, primarily due to its specific production process and impossibility of storing it. Furthermore, nowadays energy saving has become more important than ever in order to ensure its continuous supply. Predicting consumption as precisely as possible can make a significant contribution to this.

The authors of existing papers mostly deal with short-term electricity consumption prediction based on the data recorded over a short period of time (Chae et al., 2016). Unlike most previous papers, the aim of this research is to predict monthly consumption of electricity.

Predicting the electricity consumption of a single consumer is quite challenging. Predicting the electricity consumption of all consumers within a wider area is much more challenging, and requires special research approaches. A huge number of different consumers with various characteristics within the same area further complicates the prediction of their future consumption, especially at a time of increasing global consumption.

The target dataset used in this research contains the information about the actual monthly consumption of all consumers within the territory of the entire City of Užice during a period of four years. Actual problems, like this one, often involve a lot of specifics, which cannot be solved using standard, well-known procedures, and neural network (NN) architectures with simple, fully connected (dense) layers.

This paper has three main goals:

1. to present the methodology that would ensure the prediction of monthly electricity consumption in each month of the next year with the highest possible accuracy;

2. to develop and test the capacity of the NN model proposed to this aim;

3. to propose the architecture for the application which would be easy to access by end users.

The first mentioned approach involves clustering a set of data into a number of more homogeneous sets (clusters).

The second one involves developing and training NN models. After training, the developed models are simultaneously tested for the whole dataset and its subsets, which were obtained using different types of clustering. This paper presents the advantages of the proposed hybrid NN model architecture employed for predicting monthly electricity consumption.

https://doi.org/10.24846/v32i1y202307

What makes the model specific is the use of custom layers based on weight normalization, and a special type of layers which normalize the inputs across the feature maps.

This model is particularly important because it provides a possibility to predict consumption in an area including all types of consumers, which makes a dataset extremely heterogeneous. The predictive power of the proposed model is presented, together with the results provided by the model with a more classic, standard architecture, but with the same parameters.

Finally, the architecture of the ECPS, as an interactive GUI intended for the end user, is proposed. Both distribution companies and individual consumers can benefit from the results of this research because they will be able to predict their own consumption using the proposed ECPS.

In addition to the Introduction, this paper comprises five sections. Section 2 provides a brief overview of the developments in this area. Section 3 introduces four different clustering types and two types of feed-forward NN models for each cluster. Section 4 sets forth the results achieved by the models proposed in Section 3 for each clustering type. Finally, in Section 5 the conclusions of this paper are presented.

2. Literature Review

According to Walther & Weigold (2021) the prediction of electricity consumption has been drawing increasing attention of researchers over the last 25 years. Many existing papers dealing with the prediction of electricity consumption aim to predict the consumption within a given area by using different data mining techniques. Tso & Yau (2007) and Azadeh et al. (2007) compared regression analysis, decision trees and neural networks or integrating neural networks with genetic algorithms.

In the existing literature, there are authors who deal with the prediction of long-term electricity consumption. Azadeh et al. (2010) proposed the 'adaptive-network-based fuzzy inference system (ANFIS) for long-term electricity consumption prediction'. Kargar & Charsoghi (2014) used an artificial neural network model in order to predict the annual electricity consumption in Iran. Having compared two neural network models (ARIMA and Perception), the authors found out that NARX neural networks predicted electricity consumption most precisely. Yoo & Myriam (2018) noticed that the analysis of data classified by the month of the year has a great influence on the prediction of monthly expenses.

According to Hippert et al. (2001) the authors who resort to a short-time prediction, ranging between a few minutes and a few hours or days, prevail. Banga et al. (2021) compared dozens of algorithms in order to find the most accurate one for short-term forecasting. Aman et al. (2014) concluded that the best results were achieved when predicting consumption for the next few minutes. Most of the authors such as Ilić et al. (2012), Selakov et al. (2012) or Amjady & Keynia (2009) have proposed a new, hybrid short-term forecasting method. In addition to short-term electricity consumption prediction models, machine learning has been frequently used lately to predict the short-term electricity demand. Mohammadigohari (2021) confirmed that the electricity demand could be modeled using machine learning algorithms.

The following authors used machine learning techniques to predict electricity consumption within a specific area such as a city: Palermo, Italy (Beccali et al., 2008), Xinjiang, China (Song et al., 2019), Užice, Serbia (Knežević & Blagojević, 2019), a whole country: Iran (Kargar & Charsoghi, 2014), Turkey (Kavaklioglu et al., 2009) or even within a region as is the case in (Hu, 2017). On the other hand, most research is based on homogenous datasets, observing only one type of consumers in a residential building (Cai et al., 2019), public building (Zekić-Sušac et al., 2018), commercial building (Shapi et al., 2021) or a shopping mall (Kim et al., 2019). Moreover, there are studies in which the authors dealt exclusively with one type of consumers, for example household consumers (Le et al., 2019; Tso & Yau, 2007), industrial consumers (Sarswatula et al., 2022) or even with one type of electrical devices (Kumar et al., 2021).

Nevertheless, the neural network architecture proposed in most of the papers is simple, common, and based on classic, fully connected layers. In this research, a somewhat different architecture was employed, based on the use of a specific type of layers, i.e. Weight Normalization Layers (WNL), which other authors have mostly used for image processing (Salimans & Kingma, 2016; Xiang & Li, 2017). Moreover, the network architecture also included the Layer Normalization Layer (LNL), explained in (Ba et al., 2016).

Similarly to (Alanbar et al., 2020; Berriel et al., 2017; Knežević & Blagojević, 2019), the dataset used for testing the proposed model contains data about monthly electricity consumption, expressed in kWh. However, unlike most previous papers, the data used for the purpose of this paper includes a variety of consumers, who are not limited to a single commercial, residential or industrial building or neighborhood. Therefore, the performance of the proposed model was also tested on consumer clusters as in (Zekić-Sušac et al., 2018).

3. Materials and Methods

By exploring the available data, one can understand the specifics of this problem, which many scientists are trying to solve by approaching it in different ways, and using different techniques.

The methodology used in this paper comprised several main steps, graphically represented in the flow chart in Figure 1.

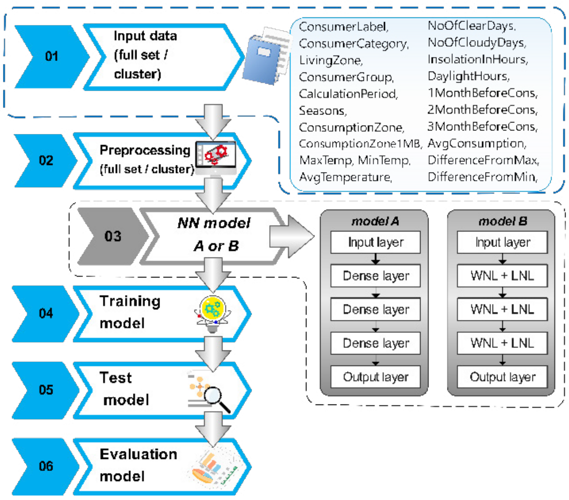

Figure 1. Research methodology flowchart

The main idea behind this research was to begin with a heterogeneous dataset, comprising monthly electricity consumption records of all existing consumers within the observed territory, and then to obtain more homogeneous datasets, convenient for the prediction.

This can be achieved by clustering the dataset according to predefined criteria. The dataset was split into multiple, smaller, more homogeneous sets, and the NN model was applied. The results obtained for each subset were compared with those obtained for the whole dataset.

To this aim, four types of clustering were performed:

- ***Clustering type I***: based on the month of the year when the consumption was recorded (twelve clusters),
- ***Clustering type II***: based on the season (four clusters),
- ***Clustering type III***: based on the consumer category (household or non-household consumers, two clusters) and
- ***Clustering type IV***: based on the consumers' living area (city or village, two clusters).

Figure 2 shows four types of consumer clustering on the territory of a city (the test system).

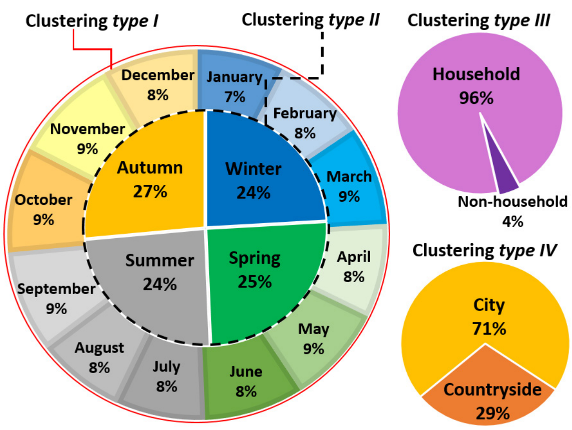

Figure 2. Different consumer clustering types in test system

The first two types of clustering were performed using the time when the consumption took place as the basis, whereas the other two were performed based on consumer characteristics (the type of consumers and their living area). After the clusterization, during the data preprocessing phase, One-hot encoding was used for all categorical attributes, except for the consumer label, for which, due to a large number of possible categories, Mean Encoding (*Target Encoding* introduced by Larionov (2020) and Pargent et al.

(2022) was usd. Efficiently encoding categorical variables is a crucial aspect during data analysis. An often encountered problem is represented by cardinality features, i.e. unordered categorical predictor variables with a high number of levels. This paper analyses techniques that yield numeric representations of categorical variables which can then be used in subsequent ML applications. The focus is on the impact of those techniques on a subsequent algorithm's predictive performance, and – if possible – deriving best practices on when to use which technique. A large-scale benchmark experiment was conducted, in the framework of which different encoding strategies together with five ML algorithms (lasso, random forest, gradient boosting, k-nearest neighbours, support vector machine), were compared. Unlike One-hot Encoding, which produces one binary feature per category, *Target Encoding* creates a new feature based on a combination of the existing feature and the target variable. This makes *Target Encoding* especially suitable for coding the consumer labels which comprise dozens of thousands of different categories. Data scaling was performed using the combination of the robust scaler and z-score standardization (*Standard Scaler*). First, the robust scaler scaled the features that were robust to outliers. It removed the median, and scaled the data within the range between *1st* quartile and *3rd* quartile (i.e. it measured this distance in terms of the *Interquartile Range* using the formula in equation (1), after which the *Standard Scaler* was applied (equation (2)).

$$ScaledVal. = \frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)} \tag{1}$$

$$ScaledVal. = \frac{x_i - mean(x)}{sd(x)} \tag{2}$$

where x_i is a data sample, Q_1 is the first quartile, Q_3 is the third quartile, *mean(x)* is the mean value of the feature and *sd(x)* is the standard deviation of *x*.

The next step involved using two types of feed-forward NN models for each cluster:

- ***NN model A*** (the basic NN model): contains three fully connected (dense) layers between input and output layers, each layer comprising the same number of neurons, and

- ***NN model B*** (custom NN model): each fully connected (dense) layer from the *model A* is replaced by a WNL layer followed by one LNL layer, as it is shown in Figure 1.

The main characteristic of the custom NN model is the use of Weight Normalization Layers (WNL), explained in (Salimans & Kingma, 2016) as "a reparameterization of the weight vectors in a neural network that decouples the length of those weight vectors from their direction". The main idea behind such reparameterization lies in batch normalization, but it does not introduce any dependency between the examples in a minibatch. In a WNL the L2 norm of the incoming weights instead of the variance is used to normalize the summed inputs to a neuron. The LNL layer normalizes the inputs across the feature maps, speeding up the training process of neural networks (Ba et al., 2016).

Most model parameters are the same for both types of NN:

- The number of neurons in each hidden layer is equal to the number of neurons in the input layer (Bengio, 2012; Weiwei & Hao, 2022). So, the number of neurons varies with different types of clustering.

- The rectified linear activation function (ReLU, equation (3)) (Hahnloser et al., 2000) was used as the activation function in each layer, except in the output layer, where the linear activation function was used (equation (4)).

$$R(x) = \max(0, x) \tag{3}$$

$$ü(\ ) = \tag{4}$$

- Mini-batch gradient descent (size=500) was used for both models.

- Adadelta (Zeiler, 2012) was used as the optimization algorithm for learning the network weight parameters.

- As for the learning rate, a predefined framework that adjusts the learning rate between epochs - Learning Rate Scheduler, with the same schedule for both models and each cluster, was used.

- To apply a penalty on the layer's kernel, a regularizer that applies both L1 and L2 regularization penalties (i.e. elastic net) was applied (L1=L2=0.5) (Zou & Hastie, 2005).

- To set the initial random weights for each layer and each bias initializer, He-normal (He et al., 2015) was used.
- Each experiment was performed twice, including one training extension, and better results were recorded.
- Both models were trained on 80% of the data, for each clustering type, and tested on the remaining 20% (test data).

The complete process of data processing and model implementation was performed in the Python programming language using the TensorFlow deep learning framework and Keras library.

The evaluation measures used in this paper to estimate the results include: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Adjusted R Square (AR^2), all expressed as a percentage (equations (5) to (8), respectively).

$$RMSE(\%) = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}}{\bar{y}} * 100 \qquad (5)$$

where y_i and $\hat{y}_i$ represent the recorded and predicted electricity consumption, $\bar{y}$ is the mean of the recorded electricity consumption, and N is the total sample size.

$$MAPE(\%) = \frac{}{N}\sum_{i=}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| * 100 \qquad (6)$$

$$MAE(\%) = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{|\hat{y}_i|} * 100 \qquad (7)$$

$$AR^2 = 1 - \frac{(1-R^2)(N-1)}{N-p-1} \qquad (8)$$

where: R^2 is the sample of R-Squared (the coefficient of determination, which measures how well the regression line approximates the actual data (equation (9))), and p is the number of independent variables.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}. \qquad (9)$$

4. Results and Discussion

The target dataset comprised numerical and categorical data about electricity consumers living in the area of the City of Užice. The data was recorded over a period of four years, and included over a million records, for about 40,000 consumers, thus covering all the existing measurement points on the above-mentioned territory.

The dataset was split into the training and test set, the training set containing data collected during the first three years, and the test set containing data from the last year which was not included in the training set, which complies with Mohammadigohari (2021). Both models were trained using 80% of the data, for each clustering type, and tested using the remaining 20% (test data). This represents a complicated though more realistic situation for the proposed model. Each experiment was performed twice, including one training extension, and the better results were recorded.

Exploring the data, it can be seen that the amount of consumption and the number of consumers had an upward trend during the observed period (Figure 3). This is an additional reason why the results obtained in this manner should be considered reliable. In Figure 3, the green line represents the number of valid measurements (the

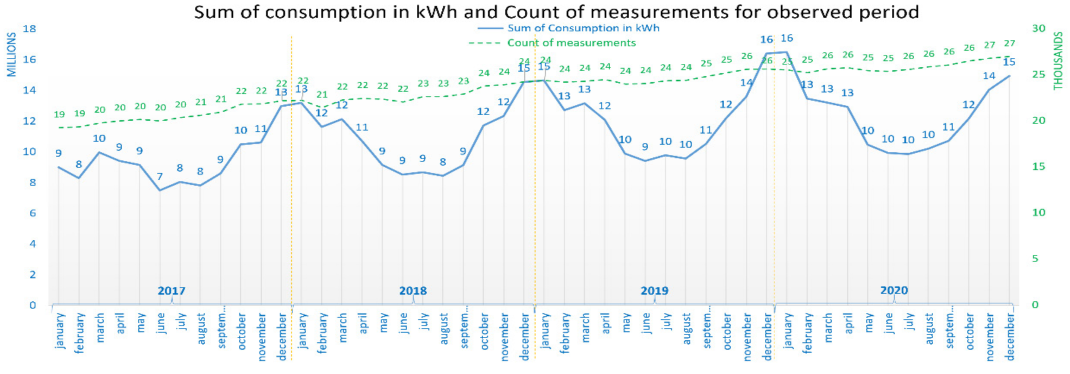

Figure 3. Sum of consumption (in millions of kWh) for all consumers over a period of four years

units relate to thousands) for each month, and the blue line represents the consumption (in millions of kWh) for these measurements.

For a certain number of months, the dataset contained irregular readings, which were removed from the final set, and that is why the diagram oscillates (i.e. it caused the line drop). This territory contains very heterogeneous consumers in relation to many criteria: category, consumer group, living zone, etc. By analyzing the dataset with regard to the living zone, two categories can be distinguished: the city – 71%, and countryside – 29%, as it is shown in Figure 2. On the other hand, on this territory, 96% of the consumers are households, whereas 4% of them are non-household consumers. The results obtained in the testing phase for each month of the last year are presented in the next subsections.

Type I clustering

Table 1 shows the prediction results obtained by NN models *A* and *B* for *type I* clustering, based on the month of the year, including those for the whole dataset (the set before clustering) in the last column.

Comparing the results obtained by both NN types, *NN model B* generally provided significantly better results than *NN model A*. The *NN model B* obtained better values for each evaluation measure, that is RMSE(%), MAPE(%), MAE(%) and AR^2(%).

Type II clustering

Table 2 shows the results obtained by NN models *A* and *B* for *type II* clustering, based on the seasons. This type of clusterization also provided better results with *NN model B* compared with *NN model A* (Table 2). On the other hand, when each season was observed as a separate dataset, the results obtained for each evaluation measure were significantly better than those obtained using the whole dataset as a unit.

Type III clustering

As for the clustering based on the consumer category (clusterization *type III)*, similarly to the previous ones, for each measure the *NN model B* provided better results than *NN model A* (Table 3 (a)). With regard to the results for subclusters, they were a bit more favourable for household consumers than for the whole dataset. For the non-household consumer category, the results were a bit less favourable. The reason for this could be the fact that this subcluster has the fewest instances (only 6% of all consumers belongs to the non-household category).

Type IV clustering

Type IV clustering, based on the zone where consumers live, provided similar results in comparison with the previous ones. For each measure in both clusters, *NN model B* had better

Table 1. Comparison between evaluation measures for *type I* clustering (per month)

MEASURE	NN TYPE	JAN	FEB	MARCH	APRIL	MAY	JUN	JULY	AUG	SEPT	OCT	NOV	DEC	WHOLE SET
RMSE (%)	A	7.89	8.48	8.06	8.61	9.8	10.12	10.11	10.24	9.33	9.1	8.17	8.39	8.34
	B	7.2	7.4	6.98	6.77	8.61	8.01	7.72	8.66	7.97	7.62	7.42	6.79	7.9
MAPE (%)	A	6.29	6.65	6.19	6.68	7.51	8.3	8.39	**8.63**	7.13	6.8	6.37	6.43	6.2
	B	5.56	5.42	5.07	**4.79**	5.87	5.81	5.75	5.23	5.74	5.5	5.09	4.83	5.68
MAE (%)	A	5.92	6.18	5.81	6.27	7.09	7.57	7.71	7.96	6.7	6.47	5.83	6.18	5.92
	B	5.3	5.22	4.84	4.74	5.77	5.65	5.57	6.63	5.58	5.33	4.86	4.72	5.46
AR^2 (%)	A	95.26	96.46	96.88	95.48	94.65	93.81	90.96	69.72	95.13	95.73	97.35	96.31	95.67
	B	96.12	97.34	97.68	97.16	96.12	96.14	94.78	98.02	96.45	97.01	97.95	97.61	96.25

Table 2. Comparison between evaluation measures for *type II* clustering (per season)

MEASURE	NN TYPE	SUMMER	AUTUMN	WINTER	SPRING
RMSE (%)	A	9.83	8.20	7.87	9.03
	B	7.42	7.75	6.92	6.80
MAPE (%)	A	**7.93**	6.20	5.79	6.97
	B	5.47	4.89	**4.81**	5.28
MAE (%)	A	7.27	5.78	5.57	6.51
	B	5.13	5.34	4.70	4.72
AR^2 (%)	A	93.42	96.67	96.60	95.12
	B	96.75	95.97	97.66	97.45

Table 3. Comparison between evaluation measures for *type III* clustering (per consumer category) and for type IV clustering (per zone)

		(a) *type III* clustering		(b) *type IV* clustering	
MEASURE	NN TYPE	NON-HOUSEHOLD	HOUSE-HOLD	CITY	COUNTRYSIDE
RMSE (%)	A	16.57	12.52	17.64	13.79
	B	9.30	7.09	6.93	7.53
MAPE (%)	A	14.85	10.74	16.63	10.78
	B	6.58	4.86	4.87	4.97
MAE (%)	A	12.29	9.26	13.26	9.84
	B	6.10	4.75	4.70	4.92
AR^2 (%)	A	63.53	70.52	91.44	91.23
	B	95.89	96.88	96.88	96.61

prediction power *than NN model A*, as it is shown in Table 3 (b).

The results presented in the previous subsections (Tables 1 to 3), with regard to both individual clusters and their comparison with the results obtained using the whole set, undoubtedly prove that the proposed model has a great prediction power in real-world conditions.

Moreover, the hyperparametres, identically used with both models, proved not to have enough influence and power to overcome the shortcomings of the classic layers in *NN model A* compared with the proposed model. In Tables 1 to 3, the values marked in green represent the best results (the minimum mean absolute error expressed as a percentage) obtained using *model B* with all clustering types. The values marked in red are the worst values for each type of clustering, obtained using *model A*. To provide a clearer picture of the results obtained using *NN models A* and *B*, Figure 4 shows the average MAPE(%) for each cluster compared with that of the whole set. Here it can be seen that significantly lower values of MAPE(%) per cluster were obtained using *NN model B*. The success of *model B* is particularly evident with clustering *types III* and *IV*, where the difference between the prediction errors is the biggest.

Generally speaking, the *NN model B* stability is significant for all clusters, which is not the case with *model A*. The value of AR^2 measure is especially important as it unambiguously indicates the stability and reliability of *model B* compared with *model A*.

With regard to the different types of clustering and the results obtained, model B is far better with clustering type III, and the value of AR2 for the non-household consumer cluster is better than the value presented by Sarswatula et al. (2022), even though they observed industrial consumers only. Even though the average MAPE(%) given in Figure 4 indicates the better characteristics of model B, the power of this model becomes

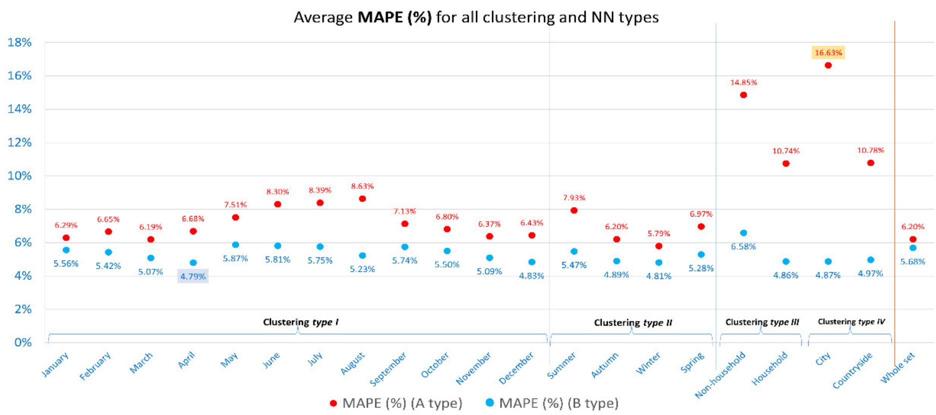

Figure 4. Review of average MAPE(%) for each cluster and whole set obtained by NN *models A* and *B*

much clearer when a single, random consumer is observed (Figure 5). The figure shows one mark (dot) for each clustering type for each month. Blue dots represent absolute errors as a percentage, obtained by Type I clustering, while the red ones represent the same measure obtained for the same consumer using the whole dataset. As it can be seen, red dots are the highest ones on the diagram, while the blue ones are on the bottom. Model B proved to be much more precise than the one with ordinary, fully connected layers, but with the same parameters, which makes it more convenient for the prediction of discrete variables even though, according to the literature the WNL type of layers (used in model B) is mostly used for image processing with Convolutional Neural Networks. These results also justify the use of neural networks for predicting electricity consumption, which is in line with the findings of Tso & Yau (2007), who concluded that neural networks outperformed various regression tools when predicting electricity consumption.

It is especially important to emphasize that the proposed model becomes even more precise when the dataset is split into more homogeneous subsets such as the four types of clustering described above, and to highlight the impact of analyzing consumption data classified by the month of the year. This complies with the results presented in (Yoo & Myriam, 2018), who divided data in a similar way, but is opposed to the conclusions drawn by Zekić-Sušac et al. (2018).

Contrary to the statement of Le et al. (2019) that electricity consumption can be predicted precisely only if one type of consumers is observed, for the purpose of this research the prediction of electricity consumption was performed on a monthly basis for all types of consumers, as in (Berriel et al., 2017), and higher precision was achieved. The mean absolute percentage error produced by the proposed model (≈5% for all consumers and ≈1% for the individual consumer) are several times lower than those produced in (Berriel et al., 2017).

Due to the specific nature of the problem itself, each case in the cited literature is specific. The research presented in each paper was conducted under unique conditions, using a unique dataset, which is not available to other authors for further exploitation, thus making the results of proposed methodologies and models impossible to compare. However, the low mean error values obtained using the model proposed in this paper call for further research in the same direction.

Figure 6 shows the rough architecture of the ECPS with the interactive GUI interface. ECPS is based on *NN model B*, used to predict electricity consumption in the future period (month) based on the input consumer characteristics. As for the end user, ECPS is very easy to use, and does not require any domain knowledge. The REST API is built using the Flask web framework and deployed using Gunicorn, a production-ready web server that is used to run the Flask application in an efficient and stable manner. Users can enter the appropriate data into the proper fields, press the submit button, and the system will display the predicted consumption in kWh.

The limitations of the presented methodology and proposed model could be related only to the data themselves. Maybe some other consumer

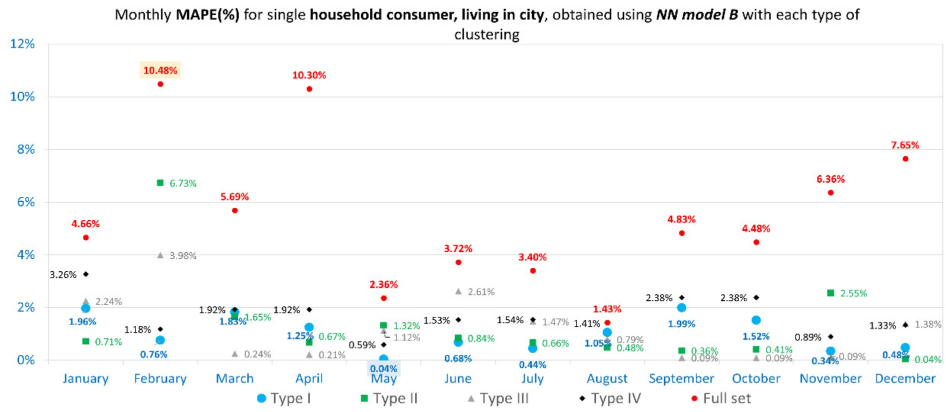

Figure 5. Review of MAPE(%) for a random consumer obtained by NN *model B* with each type of clustering

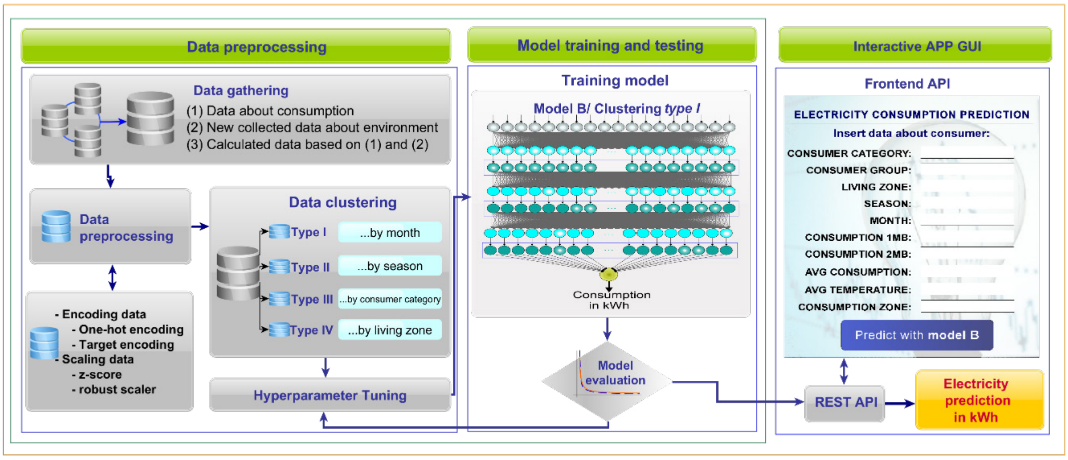

Figure 6. Rough representation of ECPS architecture

characteristics or living areas would have influenced the prediction process more than those included in the dataset and used to train the models. However, this can be easily overcome by widening the list of data recorded by distribution companies. Furthermore, the prediction results might have been even more precise if the observed period had not covered the period of the COVID-19 pandemic, which significantly affected people's lifestyles and therefore their electricity consumption.

5. Conclusion

When processing datasets that contain records of real phenomena, a lot of heterogeneous data can be encountered, with dozens of specifics that reduce the accuracy of even the most complex neural network architectures, especially when it comes to predictions based on numerical values.

In this research, two important approaches to the problem are explained, i.e. how the combination of two methodologies (in the context of splitting the chosen dataset into reasonable, homogeneous parts, and then applying the NN model with the custom architecture) can provide more precise predictions compared with predictions for the whole dataset obtained using classic, simple NN architectures. The proposed model (*model B*) ensures a substantial accuracy improvement in the prediction process for consumers which are divided into reasonable subclusters (using one of the presented clusterization types – *type I* to *type IV*).

In the worst-case scenario, the value of the mean absolute percentage error is ≈17% for the cluster obtained by *type IV* clustering and simple NN with regular, fully connected layers. Contrary to that, by applying the proposed NN architecture to the same cluster, the mean absolute percentage error reaches ≈5%, which represents a substantial improvement.

The comparison between the results for applying the proposed model (*model B*) and *model A* to subclusters and to the whole dataset clearly justifies the presented approach. When applying *model B* to an individual cluster, the mean absolute percentage error per single consumer becomes even smaller, reaching the minimum possible value of ≈0% for clustering *type I*. Therefore, the clustering *type I* and *NN model B* were proposed for the most precise prediction of electricity consumption, and they were also applied in the ECPS GUI interface.

Not only does *model B* ensure significantly lower prediction errors per cluster, but it also ensures higher outcome stability taking all the performed experiments into account, regardless of the clustering type, thus proving that the proposed application has the potential of using the trained model with new input data about the characteristics of a new consumer. Future research should focus on developing new architectures, which would be more flexible, and ensure a greater accuracy. Finally, future research should try to determine how the COVID-19 pandemic affected the electricity consumption and its prediction.

REFERENCES

Alanbar, M., Alfarraj, A. & Alghieth, M. (2020) Energy Consumption Prediction Using Deep Learning Technique. *International Journal of Interactive Mobile Technologies (IJIM)*. 14(10), 166-177. doi: 10.3991/ijim.v14i10.14383.

Aman, S., Frincu, M., Charalampos, C., Noor, U., Simmhan, Y. & Prasanna, V. (2014) *Empirical comparison of prediction methods for electricity consumption forecasting*. Technical Report. https://tinyurl.com/3cmeejbh [Accessed: 15 June 2022].

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

Azadeh, A., Ghaderi, S. F., Tarverdian, S. & Saberi, M. (2007) Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. *Applied Mathematics and Computation*. 186(2), 1731–1741. doi: 10.1016/j.amc.2006.08.093.

Azadeh, A., Saberi, M., Nadimi, V., Iman, M. & Behrooznia, A. (2010) An integrated intelligent neuro-fuzzy algorithm for long-term electricity consumption: Cases of selected EU countries. *Acta Polytechnica Hungarica*. *7*(4), 71-90.

Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016) Layer normalization. To be published in *Machine Learning (stat.ML).* [Preprint] https://arxiv.org/abs/1607.06450 [Accessed: 20 June 2022].

Banga, A., Ahuja, R., & Sharma, S. (2021) Stacking machine learning models to forecast hourly and daily electricity consumption of household using Internet of Things. *Journal of Scientific & Industrial Research*. 80, 894-904.

Beccali, M., Cellura, M., Brano, V. L. & Marvuglia, A. (2008) Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area. *Renewable and Sustainable Energy Reviews*. 12(8), 2040-2065. doi: 10.1016/j.rser.2007.04.010.

Bengio, Y. (2012) Practical recommendations for gradient-based training of deep architectures. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, (2nd ed.) vol. 7700, pp. 437-478. Springer.

Berriel, R. F., Lopes, A. T., Rodrigues, A., Varejao, F. M. & Oliveira-Santos, T. (2017) Monthly energy consumption forecast: A deep learning approach. In: *2017 Proceedings of the International Joint Conference on Neural Networks*, *May 14-17, 2017, Anchorage, Alaska, USA*. Institute of Electrical and Electronics Engineers (IEEE). pp. 4283-4290.

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

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

Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J. & Seung, H. S. (2000) Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*. 405(6789), 947-951.

He, K., Zhang, X., Ren, S., & Sun, J. (2015) Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE International Conference on Computer Vision*, *ICCV `15, December 7-13, 2015, Santiago, Chile*. *Washington, DC, USA.* IEEE Computer Society. pp. 1026-1034.

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

Hu, Y.-C. (2017) Electricity consumption prediction using a neural-network-based grey forecasting approach. *Journal of the Operational Research Society*. 68(10), 1259-1264. doi: https://doi.org/10.1057/s41274-016-0150-y.

Ilić, S. A., Vukmirović, S. M., Erdeljan, A. M. & Kulić, F. J. (2012) Hybrid artificial neural network system for short-term load forecasting. *Thermal Science*. *16*(suppl. 1), 215-224. doi: https://doi.org/10.2298/TSCI120130073I.

Kargar, M. J. & Charsoghi, D. (2014) Predicting annual electricity consumption in Iran using artificial neural networks (NARX). *Indian Journal of Science and Research*. 5(1), 231–242.

Kavaklioglu, K., Ceylan, H., Ozturk, H. K. & Canyurt, O. E. (2009) Modeling and prediction of Turkey's electricity consumption using artificial neural networks. *Energy Conversion and Management*. *50*(11), 2719–2727.

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

Knežević, D. & Blagojević, M. (2019) Classification of electricity consumers using artificial neural networks. *Facta Universitatis, Series: Electronics and Energetics*. 32(4), 529-538. doi: https://doi.org/10.2298/FUEE1904529K.

Kumar, M., Shenbagaraman, V., Shaw, R. N. & Ghosh, A. (2021) *Predictive data analysis for energy management of a smart factory leading to sustainability*. Innovations in electrical and electronic engineering. Singapore, Springer.

Larionov, M. (2020) Sampling techniques in bayesian target encoding. To be published in *Machine Learning (stat.ML).* [Preprint] https://arxiv.org/abs/2006.01317 [Accessed: 20th June 2022].

Le, T., Vo, M. T., Vo, B., Hwang, E., Rho, S. & Baik, S. W. (2019) Improving electric energy consumption prediction using CNN and Bi-LSTM. *Applied Sciences*. 9(20), 4237.

Mohammadigohari, M. (2021) *Energy consumption forecasting using machine learning*. Master's Thesis, Rochester Institute of Technology, Rochester, NY, USA.

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

Salimans, T. & Kingma, D. P. (2016) Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In: *Advances in Neural Information Processing Systems 30: Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, December 5-10, 2016, Barcelona, Spain. Red Hook, NY, USA,* Curran Associates Inc.. pp. 901-909.

Sarswatula, S. A., Pugh, T. & Prabhu, V. (2022) Modeling Energy Consumption using Machine Learning. *Frontiers in Manufacturing Technology.* 2. doi: https://doi.org/10.3389/fmtec.2022.855208.

Selakov, A., Cvijetinović, D. & Ilić, S. (2012) Short term load forecasting using support vector machines for different consumer types. *Journal on Processing and Energy in Agriculture*. 16(3), 128-131.

Shapi, M. K. M., Ramli, N. A. & Awalin, L. J. (2021) Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. *Developments in the Built Environment*. 5, 100037.

Song, X., Liang, G., Li, C. & Chen, W. (2019) Electricity consumption prediction for Xinjiang electric energy replacement. *Mathematical Problems in Engineering*. 2019, 1–11. doi: https://doi.org/10.1155/2019/3262591.

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

Walther, J. & Weigold, M. (2021) A systematic review on predicting and forecasting the electrical energy consumption in the manufacturing industry. *Energies*. 14(4), 968. doi: https://doi.org/10.3390/en14040968.

Weiwei, W. & Hao, D. (2022) The Application of the Generalized Regression Neural Network Model Based on Information Granulation for Short-Term Temperature Prediction. *Studies in Informatics and Control*. 31(3), 53-62.

Xiang, S. & Li, H. (2017) On the effects of batch and weight normalization in generative adversarial networks. To be published in *Computer Vision and Pattern Recognition.* [Preprint] *https://arxiv.org/abs/1704.03971* [Accessed: 26th June 2022].

Yoo, S. G. & Myriam, H.-Á. (2018) Predicting residential electricity consumption using neural networks: A case study. *Journal of Physics: Conference Series*. 1072(1), 012005. doi: https://doi.org/10.1088/1742-6596/1072/1/012005.

Zeiler, M. D. (2012) Adadelta: An adaptive learning rate method. To be published in *Machine Learning (stat.ML).* [Preprint] https://arxiv.org/abs/1212.5701 [Accessed: 20th June 2022].

Zekić-Sušac, M., Scitovski, R. & Has, A. (2018) Cluster analysis and artificial neural networks in predicting energy efficiency of public buildings as a cost-saving approach. *Croatian Review of Economic, Business and Social Statistics*. 4(2), 57-66. doi: https://doi.org/10.2478/crebss-2018-0013.

Zou, H. & Hastie, T. (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301-320. doi: https://doi.org/10.1111/j.1467-9868.2005.00503.x.