DIGITALES ARCHIV

Buturache, Adrian-Nicolae; Stancu, Stelian

Article

Building energy consumption prediction using neural-based models

Provided in Cooperation with:
International Journal of Energy Economics and Policy (IJEEP)

This Version is available at:
http://hdl.handle.net/11159/8609

Building Energy Consumption Prediction Using Neural-Based Models

Adrian-Nicolae Buțurache[1]*, Stelian Stancu[2]

[1]Economic Cybernetics and Statistics Doctoral School, Bucharest University of Economic Studies, Bucharest, Romania,
[2]Department of Informatics and Economic Cybernetics, Bucharest University of Economic Studies, Bucharest, Romania.
*Email: ad.buturache@yahoo.ro

ABSTRACT

In the recent years digital transformation became one of the most used approaches in building energy consumption optimization. Increased interest in improving energy sustainability and comfort inside buildings has created an opportunity for digital transformation to build predictive tools for energy consumption. By retrofitting or implementing new construction technologies nowadays the quantity and quality of the operational data collected has reached unprecedented levels. This data must be consumed by implementing powerful predictive tools that will provide the needed level of certainty. Adopting six sigma's define, measure, analyze, improve, control (DMAIC) cycle as predictive analytics framework will make this paper accessible for both professionals working in energy industry and researchers that are developing models, creating the premises for reducing the gap between research and real-world business, guiding the use of data. Moreover, the selected strategy for preprocessing and hyperparameter selection is presented, the final selected models showing scalability and flexibility. At the end the architectures, performance and training time are discussed and then coupled with the thought process providing a way to weigh up the options. Building energy consumption prediction, it is a relevant and actual topic. Firstly, on European level, meeting the targets set by the new European Green Deal for buildings sector is relying heavily on digitization and therefore on predictive analytics. Secondly, on Romania level, the liberalization of the energy market created an unpreceded energy price increase. The negative social impact might be diminished not only by the price reduction, but also by understanding how the energy is consumed.

Keywords: Machine Learning, Artificial Neural Networks, Building Energy Prediction, Six Sigma
JEL Classifications: O13, O14, O31, Q47, C45, R11

1. INTRODUCTION

Energy consumption prediction represents one of the main concerns of the modern world. Since the Industrial Revolution, energy consumption has gained another dimension. Our lifestyles and energy consumption habits are increasingly interdependent, encompassing demands for electricity, steam, or hot or chilled water. The cost of energy is increased by the environmental costs associated with the pollution generated by the entire conversion process, from raw resources to refined end-user products. Prediction models are essential in energy management and planning. Buildings' energy consumption can be improved

in three ways: system improvement, device improvement, and behavior improvement. System and device improvement are closely related to technological advances while behavior is driven by education and awareness. All three development directions should be guided by coherent laws and regulations that are eventually aligned on a global scale. Going forward, two main approaches to potentially improve construction have been identified: new buildings should be more efficiently designed than existing buildings and existing buildings should be retrofitted to reduce energy consumption. Traditional grid solutions are limited to electrical power distribution, while smart grids represent an evolution of the traditional grid, enabling a two-way interaction

between suppliers and customers (Vrablecová et al., 2018). The main aim is to optimize in real time how energy is delivered to customers. A smart grid must monitor, learn, predict, and drive actions. One of the major challenges in renewable energy is to sustain the continuous supply of power from various production sites in the desired quantity when required. Smart grid solutions play a key role in the successful deployment of green energy technologies. Energy management systems are responsible for cost minimization and quality optimization and are crucial for smart grid operations. In the last 170 years, the growth rate of global energy consumption has been around 2.4% per year, with no indications that it will decrease (Jarvis et al., 2012). During the last 50 years, events such as the rise in oil prices in the 1970s, nuclear accidents, global climate change, renewable and sustainable energy technology breakthroughs, and the rapid growth of emerging economies through industrialization have made energy consumption reduction an important research field, particularly since the demand shows no signs of slowing down (U.S. Energy Information Administration, 2019). The International Energy Agency (IEA) reports that buildings represent the largest energy-consuming sector and that this sector continues to increase annually (International Energy Agency, 2013, 2018; U.S. Energy Information Administration, 2020). Models for predicting energy consumption are divided into two major classifications: based on model type and based on model prediction time horizon. Model type can be physical, statistical, machine learning or hybrid, while the model prediction time horizon can be short, medium, or long. The major advantage of machine learning and statistical models is their flexibility. Physical modeling can be detailed and precise, but at the same time, it is only appropriate for specific use cases (Reimann et al., 2018). Building demand for electricity depends on various parameters of the building itself, such as glazing percentage and properties, building fabrics, occupancy pattern, number of floors, level of internal gains, and building purpose (Korolija et al., 2013). Energy consumption prediction for the medium and long term represents one of the core information sources for strategic and tactical decisions concerning areas such as development directions, capital investment, revenue analysis, or capacity management.

The new European Green Deal is targeting improvements in the way the energy is consumed in the buildings sector. The Council of European Union highlights two important areas: renovation of the existing buildings to increase their efficiency and eco-driven design for the new buildings to be built. European Union Renovation Wave and Innovation Fund represents two frameworks made to enable professionals to tackle the challenges on the way to a decarbonized Europe. For the already existing buildings and for the new to be built these two frameworks provide the tools, including regulations and financing, to optimize and decrease energy consumption. On top of them are the NextGenerationEU and InvestEU funds that acts like a binder since are aiming areas like: lead on energy efficient artificial intelligence solutions, data share across EU, usage of technologies to make buildings more energy efficient. The access on energy must be viewed from three different perspectives. First one is related to the infrastructure and its existence, the second is related to the capability of producing the quantity of energy needed and the third is related to the

possibility of the end customer to buy the energy. The access on electricity and adoption of the latest technologies is a measure of wellbeing. A healthy and sustainable development will reduce the gaps between different social classes. Through digitization and predictive analytics, the existing and limited resources available can be shared in a more even and cheap way, both acting like enablers.

Up to 90% of the total energy used during a building's life cycle is used during building operations. Of this percentage, up to 20% could be saved through the adoption of a proactive attitude toward energy control and fault detection (Ramesh et al., 2010; Teke and Timur, 2014), in other words by introduction of predictability. This article is providing an overview on the use of neural-based model on predictive analytics of building energy consumption. The focus is on modeling, by identifying and highlighting the theoretical and practical considerations of neural-based algorithms for building energy consumption prediction. In the end, the outcome of the best performing models are compared in terms of resources spent for training and generalization capacity.

Building energy consumption optimization represents a relevant and actual topic considering all the initiatives started at European level. The guidelines for the upcoming 10 years are clear, and the premises are that this topic will remain relevant and actual at least until 2050.

2. LITERATURE REVIEW

Artificial neural networks (ANNs) represent one of the most used machine learning models for energy consumption prediction (Amasyali and El-Gohary, 2018). Data analysis on a house built to testing new technologies for improving energy efficiency, indoor air quality, and sustainable construction highlighted the benefits of a straightforward approach (Biswas et al., 2016). FFNN models were used to predict energy consumption based on weather data gathered over 3 months. Energy consumption and HVAC equipment data were recorded with a five-minute timestamp. During the data collection process, the house was unoccupied, and the impact of the occupant's behavior was not included in the data. Two ANN models based on Levenberg-Marquardt and OWO-Newton algorithms were deployed to predict total energy consumption. Input data and network topology were the same for both models: three neurons in the input layer, seven neurons in the hidden layer, and one neuron in the output layer. This simple model architecture proved sufficiently powerful to predict energy consumption with coefficients of determination between 0.87 and 0.91. Comparison of multiple machine learning techniques showed that ANN can perform better than linear regression and support vector machines on long-term prediction. In total, 4 years of data including independent variables such as ambient temperature, installed power capacity, resident electricity consumption, and gross domestic product were used (Ekonomou, 2010). In the same long-term energy consumption paradigm and the same types of models, ANN, linear regression, and least-squares SVM were compared using gross electricity generation, installed capacity, total subscribership, and population as independent variables (Ekonomou, 2010). Another comparison between machine learning

techniques—this time for short-term prediction, with 15-minute resolution data for day-ahead prediction—proved the superiority of ANN over other techniques, such as linear regression, support vector machine, RBF kernel, and nearest neighbor ball tree (Chae et al., 2016). A deep learning approach can be tried using LSTM algorithms (Marino et al., 2016). Standard LSTM- and LSTM-based sequence-to-sequence (S2S) architectures were tested on two benchmark datasets, the first with a resolution of 1 h and the second with a resolution of one minute. Both datasets were gathered from a single residential customer. S2S architecture performed better on both datasets, while standard LSTM architecture was unable to forecast accurately on the dataset with one-minute resolution data. A more extensive analysis was made using LSTM (Sülo et al., 2019). A Bayesian regularization neural network approach was proposed as a simplified approach for predicting a commercial building's energy consumption (Kim et al., 2019). Since the data quantity is limited, the authors predict that overfitting is likely to occur. Another simplified approach is sensitivity analysis, which has also proven useful in reducing the number of independent variables used in the analysis. Comparison of ARIMA, FFNN, DNN, conventional recurrent neural networks (CRNNs), and LSTM for short-, medium-, and long-term prediction revealed that ARIMA, CRNN, and LSTM are close in terms of performance for short-term predictions, while for medium- and long-term predictions, RNN and LSTM outperformed all other models (Nugaliyadde et al., 2019). To overcome the varying nature of renewable energy sources, an artificial neural network-based predictive model for optimizing and energy usage schedules can reduce the effects experienced by customers (Finck et al., 2019). Moreover, compared with a conventional approach, such as the proportional-integral controller, selected flexibility indicators are improved.

There is no single best algorithm for energy consumption prediction exists. However, a review of the existing work led strong expectations for the selected path. Although FFNN and LSTM are completely different models, they still appear in many comparatives analyzes. Neural-based modeling approach enable researchers to study building energy consumption without having priori experience in this field. Being scalable and flexible these models outperform existing models. Furthermore, the key is to understand neural-based modeling fundamentals, industry needs and, in the end, to refine the models to meet the professionals' expectations. On a macro level the enablers for these neural-based solutions and any other type of data analytics are all the frameworks proposed on European level where the regulations and fundings are driving digital transformation. On a micro level, the incremental adoption of these solutions will depend on the quality of the results delivered.

3. THEORETICAL FUNDAMENTALS

3.1. Feed Forward Neural Networks

ANNs have the advantage of providing robustness for non-linear problems and offer the possibility of scaling the solution. ANN represents a mathematical model of the human nervous system (Kumar et al., 2013). FFNNs consist of simple calculation units called "neurons" operating in parallel. Neurons are organized in

layers. Each layer can contain one or more neurons. The input layer the same as the output layer represents the only two areas of the network in which interaction with the outside environment is possible. The input layer is used to feed the network with data. The output layer contains the predictions made by the network. Consecutive layers are connected, and each connection has a synaptic weight attached. Synaptic weights express the importance of a given input at a given time (Figure 1).

The learning algorithm represents the procedure whereby the synaptic weights are adjusted to minimize the objective function (Figure 2). The synaptic weights can be said to store the knowledge. Under the supervised learning paradigm, predicted values are compared with real values during the training process. Based on the resulting error, all synaptic weights are updated. The function used to determine the difference between actual and predicted values is called the "cost" or "loss".

Inside the neuron, u is calculated—the sum of the dot product of every pair as in equation (1).

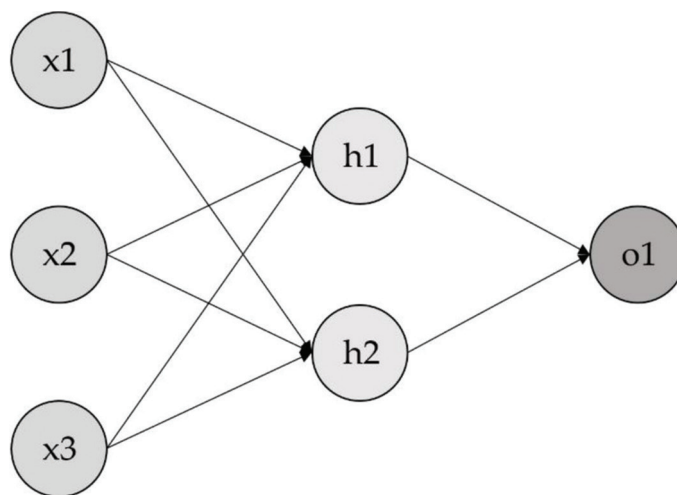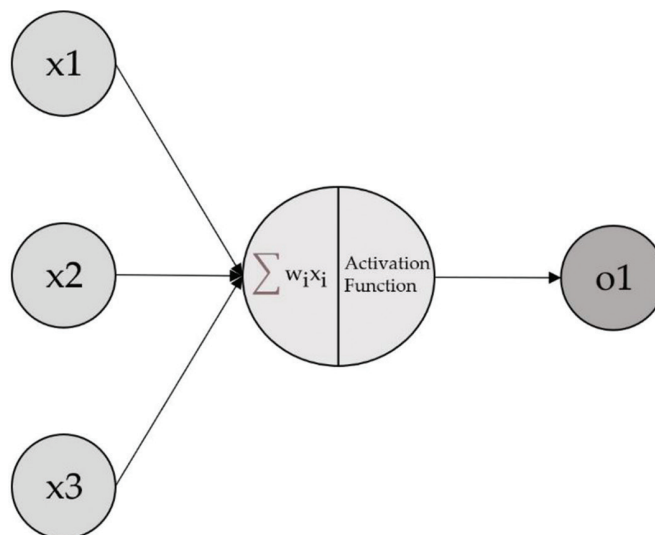**Figure 1:** Feed forward neural network schema

Figure 2: Artificial neuron mathematical abstraction

$$u = w_1 x_1 + w_2 x_2 + w_3 x_3 + b \qquad (1)$$

On the summation of the dot product can be added a bias, b, necessary to add robustness and to avoid getting blocked in a local minimum. The activation function is used to trigger or not trigger the neuron once the weighted sum of the inputs exceeds a limit. Non-linearity is thus introduced into the neuron output. This feature is important in actual case scenarios since most of the studied problems rely on non-linear data. Going forward with the logic, a non-linear model can build non-linear decision boundaries, which will lead to a better model fitting. The rectified linear unit (RELU) activation function performs the following operation:

$$f(x) = \max(0, x) \qquad (2)$$

By its nature, it is more computationally efficient than sigmoid or tanh activation functions. RELU overcomes the vanishing gradient problem, enabling speed and performance (Glorot et al., 2010). Rectified linear units have become popular among machine learning practitioners, with convolutional neural networks used simultaneously for image recognition (Kusuma and Afiahayati, 2018). The output of the neuron, *Y* is calculated by passing the summation of the dot product through the activation function. The result can or cannot activate the neuron, based on the threshold.

$$Y = f(u) = f(w_1 x_1 + w_2 x_2 + w_3 x_3 + b) \qquad (3)$$

3.2. Feed Forward Neural Networks

Recurrent neural networks are neural networks designed for sequential data and predict the next step of the sequence with respect to the sequence's previous steps. CRNNs are discrete-time dynamical systems that possess an input, an output, and a hidden layer (Pascanu et al., 2013). One of their main limitations is attributable to the vanishing and exploding gradients (Bengio et al., 1994). The synaptic weight that connects hidden layers of consecutive states (i.e. t-3, t-2, t-1, t, where t is the current state) is the same. If it is too small, the gradient becomes increasingly lower until it vanishes. If it is too large, the gradient becomes increasingly larger until it explodes. This is an effect of the training conducted with gradient-descent based algorithms and computations completed by backpropagation through time (BPTT) (Werbos, 1990). BPTT is similar to the backpropagation (Rumelhart et al., 1986) used for FFNNs. The main difference is that the gradient is calculated individually for each time step of the RNN, and at the end, the resulting gradients are added. Another weakness is due to the information morphing, which reveals the network's inability to maintain relevant information if the analyzed context contains several time steps (i.e., relevant information occurring at time step t-15 may be lost until the current state t is calculated). To surpass the issues related to the conventional RNN, another type of gradient-based method called LSTM was introduced by Hochreiter and Schmidhuber (1997) (Figure 3). This solution proposes adding gating functions to the state dynamics. These functions enable the network's ability to remember information from the earlier stages. LSTM is equipped with three gates: the input, output, and forget gates. Compared with the conventional recurrent neural network, which has only one neural network in each cell, LSTM has four. The cell's gates—input, output, and

Figure 3: LSTM cell
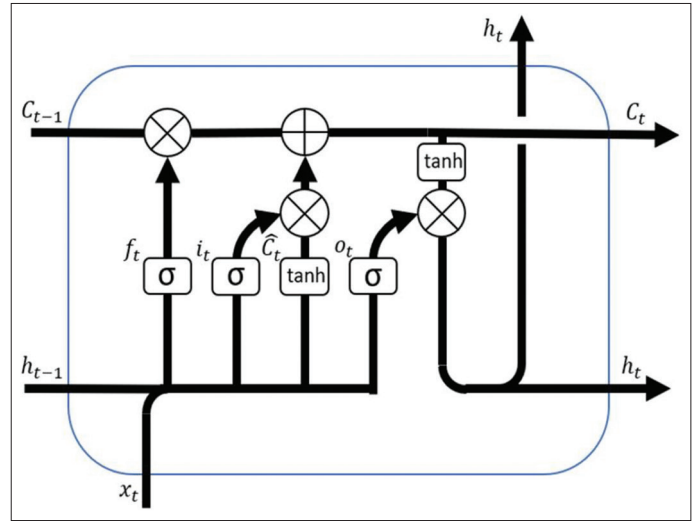
forget—determine which information is passed or blocked and are composed of the neural networks mentioned above. All three gates possess sigmoid-activated neural networks with outputs of 0 and 1.

$$\text{sigmoid}(t) = \frac{1}{1 + e^{-t}} \qquad (4)$$

$$\tanh(t) = \frac{e^z - e^{-t}}{e^t + e^{-t}} \qquad (5)$$

where t represents the current state. When the value in the gate is 0, the information passing is blocked; when the value in the gate is 1, the information can pass through the gate in its entirety.

Long-term memory cells are described using the following equations:

$$f_t = \sigma(x_t U^f + h_{t-1} W^f) \qquad (6)$$

$$i_t = \sigma(x_t U^i + h_{t-1} W^i) \qquad (7)$$

$$\widehat{C}_t = \tanh\left(x_t U^g + h_{t-1} W^g\right) \qquad (8)$$

$$C_t = \tilde{A}\left(f_t C_{t-1} + i_t \widehat{C}_t\right) \qquad (9)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o) \qquad (10)$$

$$h_t = \tanh(C_t)\, i_t \qquad (11)$$

The first equation is the forget gate and is used to decide how much of the information will be ignored and how much will be stored in the current cell state. Equations (7) and (8) control how much of the new computed information will be written in the cell state C_t. $\widehat{C}_t$ is the vector of the new candidate values for the current state cell. The current state calculation is a function of the previous cell state, C_{t-1}, multiplied by (taking the decision on what to forget from the previous state). To this is added $\widehat{C}_t$ but only after multiplication with the input gate. This multiplication basically allows only a certain amount of the input information to be part of the current state (9). At the end, the output is composed

of (10) and (11), representing the output gate and the hidden state output. The output gate determines what information is used for prediction and determines what information is sent to the next layer.

Both FFNNs and LSTMs can be trained using gradient-descent-based algorithms. More generally, gradient-descent-based algorithms are used to find the local minimum or maximum of a differentiable function. To search for the maximum, the steps taken to find the solution are proportional to the gradient. To search for the minimum, the steps taken are opposite to the gradient. Adam, derived from adaptive moment estimation, is a method for stochastic optimization based on adaptive estimates of low-order moments for first-order gradient-based optimization of stochastic objective functions (Kingma and Ba, 2014). For a given objective function $J(\Theta)$ parametrized by the model's parameters Θ, the update equation can be written as follows:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\widehat{v_t}} + \varepsilon} \widehat{m_t} \tag{12}$$

$$\widehat{m_t} = \frac{m_t}{1 - \beta_1^t} \tag{13}$$

$$\widehat{v_t} = \frac{v_t}{1 - \beta_2^t} \tag{14}$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{15}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{16}$$

where $\widehat{m_t}$ = compute bias-corrected first moment estimate, $\widehat{v_t}$ = compute bias-corrected first moment estimate, = update biased first moment estimate, v_t = update biased second raw moment estimate, and β_1^t and β_2^t = exponential decay rates for the moment estimates at time step t and t-time step. Both FFNNs and LSTM are supervised learning algorithms. Supervised learning is achieved under the certainty that the target is known and can be split into two main sub-classes: classification and regression (Fawcett and Provost, 2013). After preprocessing all data sets, the retained independent variables had targets associated represented by the energy consumption. Energy consumption is a numeric and continuous variable, meaning that all models prepared are built to be part of the regression sub-class.

4. METHODOLOGY

The study's entire methodology was aligned with the DMAIC framework (Figure 4). DMAIC is the initialism for Define, Measure, Analyze, Improve, Control and is Six Sigma's process improvement methodology, ensuring quantifiable and sustainable results. Between these five phases, feedback loops are set to ensure that project results meet business needs and that expectations are realistic (Beemaraj and Prasath, 2013). Usually, in real-world applications, Six Sigma-based projects become the mandatory step between baseline and improved operations. During the Define phase, the problem statement and the project's goal are defined. During the Measure phase, the issues related to data quality and quantity were assessed. For example, information on the number of floors was missing in proportion of 75.5%. If this study's purpose had been to predict the energy consumption at floor level for each building, this would have been impossible. Most of the time was spent on understanding and preparing the data. If the data quantity and quality are not suited to the project's scope, then the project must be stopped, its scope adjusted, or the project should be continued without the scope being refined but at a high risk of culminating in no meaningful insights.

The data used in this study were made available by ASHRE through a competition carried out on kaggle.com (ASHRAE, 2019). The scope of this study is to predict the energy consumption of 1430 buildings clustered in 16 sites. All buildings are labeled based on their primary use: education, lodging/residential, office, entertainment/public assembly, other, retail, parking, public services, warehouse/storage, food sales and service, religious worship, healthcare, utility, technology/science, manufacturing/industrial, and services (Table 1).

As part of the data preprocessing step, five of the seventeen primary use categories were retained while the remaining twelve were merged under the other category. The primary use categories ultimately used

Table 1: Data available by site and primary use (thousands)

Site ID	Education	Entertainment/public assembly	Lodging/residential	Office	Public services	Other	Total per site
0	258.5	43.6	237.2	203.7	NA	165.5	908.4
1	192.7	8.7	87.6	140.2	17.5	NA	446.9
2	535.3	183.8	105.3	210.6	52.684	96.488	1184.3
3	787.7	385.7	96.3	200.1	741.9	157.1	2369.1
4	557.3	62.1	29.6	NA	50.5	46.8	746.6
5	428.9	157.5	8.7	96.3	43.7	43.7	779.1
6	113.8	26.3	96.4	69.9	8.7	NA	315.3
7	102.7	NA	NA	NA	NA	NA	102.7
8	NA	196.2	NA	55.6	227.9	88.1	567.9
9	551.9	148.2	166.4	140.2	17.5	43.8	1068.2
10	109.4	34.5	26.2	37.3	NA	28.7	236.5
11	42.6	NA	NA	NA	NA	NA	42.6
12	175.1	17.2	NA	78.7	8.7	35.1	314.8
13	201.8	52.5	87.5	613.5	43.9	237.1	1236.5
14	227.1	87.5	78.9	330.1	61.1	105.2	890.0
15	293.2	101.8	199.6	126.8	43.1	43.1	807.6

in the study were: education, entertainment/public assembly, lodging/residential, office, public services and other (Table 2).

The available data are categorized according to source into three categories: weather data, building data, and energy consumption data. Weather data are gathered from each site's location and consist of the following variables: Site ID, timestamp, air temperature, cloud coverage, precipitation depth, dew temperature, sea level pressure, wind direction, and wind speed (Table 3).

In terms of missing data, six out of thirteen independent variables required data imputation. Removal of the records with missing data was not an option owing to the limited quantity of data. Missing data can be caused by faults in data acquisition, errors in measurement, insufficient resolution of data sampling, and lack of data acquisition hardware. The computational method used for handling missing data is the nearest neighbor. The nearest neighbor is a univariate imputation schema and relies on the start and end points of the gaps within the data to estimate what is in between. This method was selected from among two other methods—linear and cubic spline interpolation. For the data used during this research, the nearest neighbor performed better.

Relying on the theoretical foundations and existing research papers, two models were identified as suitable for energy consumption prediction: FFNN and LSTM (Figure 5).

The selection of the hyperparameters had two dimensions. The first was represented by the model's ability to generalize and the second by the speed of training and testing. Based on the mathematics behind ANNs, categorical data cannot be used in any format other than numeric data. During the preprocessing step, categorical data were transformed into numerical data through encoding. From the timestamp data, date and time information were extracted: Year, month, day, hour, weekend, working days, working hours. At the end of the preprocessing step, a data set consisting of 68 variables, including dependent variables, was obtained (Figure 6).

Selection of the number of neurons in the hidden layer and the use of a single hidden layer were part of an optimization process aimed at finding the balance between the model's ability to generalize and the time required for training. Data modeling and final architecture selection are part of DMAIC's measure and improve phases. The optimization process can be visualized as a feedback loop between these two phases.

The three metrics used for evaluation are: mean absolute error (MAE), coefficient of determination (R^2) and training time (TT). The average of the absolute error, equation (17), is used for comparing different models on the same dataset, while R^2, equation (18), it is a measure of how well the model can explain the variability in the output, this metric making the research eligible for comparisons with other models since R^2 does not depend on the dataset used.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| \tag{17}$$

Table 2: Count of building by site and primary use

Site ID	Education	Entertainment/public assembly	Lodging/residential	Office	Public services	Other	Total per site
0	30	5	27	24	19	0	105
1	22	1	19	16	0	2	60
2	61	21	12	24	11	6	135
3	92	44	11	23	18	85	273
4	66	9	4	0	6	6	91
5	49	18	1	11	5	5	89
6	13	3	11	8	0	1	36
7	12	0	0	0	0	0	12
8	0	24	0	7	11	28	70
9	63	17	19	16	5	2	122
10	14	4	3	5	4	0	30
11	5	0	0	0	0	0	5
12	29	2	0	9	4	1	45
13	23	6	10	70	27	5	141
14	26	10	9	38	12	7	102
15	41	15	28	18	6	6	114

Table 3: Weather data summary

	Air temp.	Cloud coverage	Precip. depth	Dew temp.	Sea level pressure	Wind dir.
Count	139773	139770	139773	139773	139773	139773
Mean	14.4	2.9	0.7	7.3	1016.2	179.3
Std	10.6	3.0	6.8	9.8	7.4	111.8
Min	-28.9	0	0	-35	968.2	0
25%	7.2	0	0	0.6	1012.1	80
50%	15	2	0	8.3	1016.4	190
75%	22.2	6	0	14.4	1020.4	280
Max	47.2	9	343	26.1	1045.5	360
Missing data before imputation (%)	0%	49%	0%	36%	8%	4%

Count=Total number of data points; std=Standard deviation; min=Minimum value; 25%=First quartile; 50%=Median value; 75%=Third quartile; max=Maximum value

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad (18)$$

Where $\hat{y}$ and y_i represents the predicted value and actual value.

Training time it is a measure of the resources spent on training. Training time depends on the hardware and software setup used. For this research the setup consists of a Dell Precision 7350 equipped with Intel Core I5-8400H @ 2.5 GHz CPU, 32 GB RAM, Nvidia Quadro P2000 GPU, Windows 10, and Python 3.6.10. All the algorithms were implemented using Keras and TensorFlow on the backend.

5. RESULTS AND DISCUSSION

FFNN architecture allows training that is twice as fast with an average of 2188 s and R^2 equal to 0.8424 over all primary use categories, while LSTM's average training time is 4402 s and average R^2 is equal to 0.8461 (Figure 7).

Due to their configurability and scalability neural-based models possess the capability of learning and generalizing from different datasets having different patterns. The mean absolute error, coefficient of determination and training time depend on the
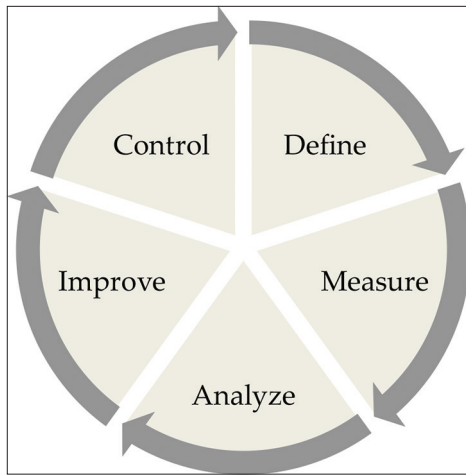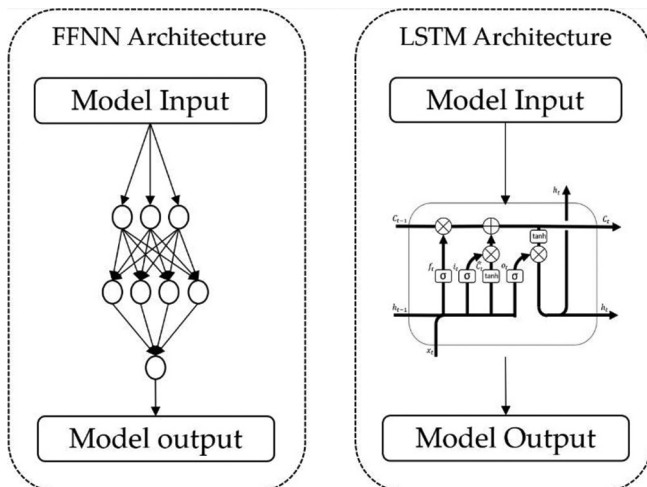
Figure 4: DMAIC methodology

Figure 5: FFNN and LSTM schema

primary use and site. This may also be linked to how the data are gathered and the extent to which the collected data can explain the phenomena. When the method for dealing with the missing data was selected, the impact of removing the rows or columns containing empty records was assessed. Data removal was not an option since these two approaches led to inferior results, while data imputation provided better results. Given that LSTM is a type of neural network built to model time series, the available optimizations are more generous for problems requiring prediction of energy consumption than in the case of FFNN. FFNN, on the other hand, excels in speed compared to LSTM, being least sophisticated. However, LSTM's mechanisms in place for capturing the short and long-term dependencies by default is require more training time (Figure 8). Given that

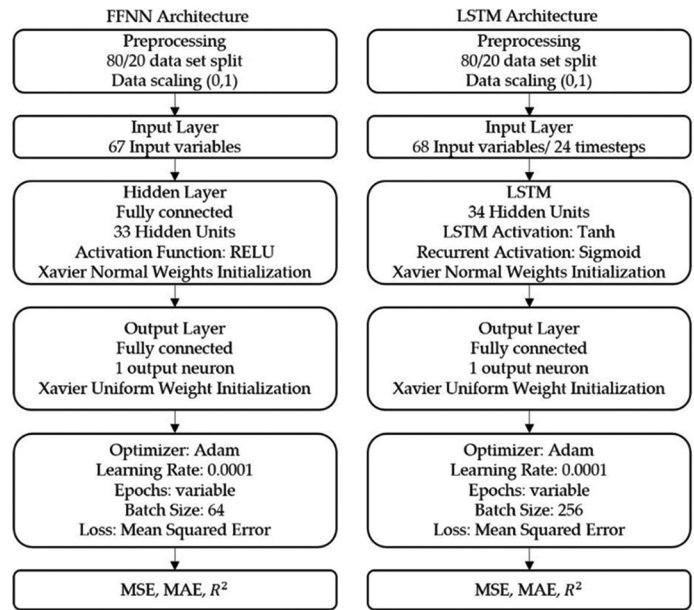**Figure 6:** Final FFNN and LSTM architectures used for prediction

Figure 7: Average R^2 comparison by primary use
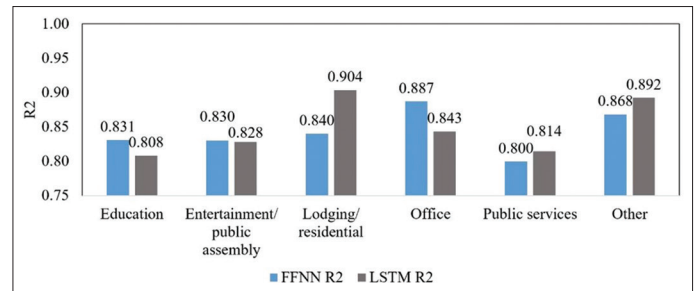

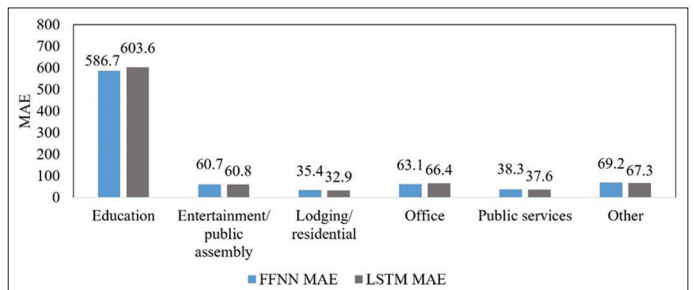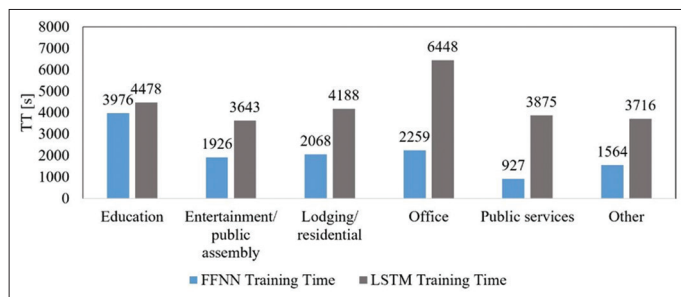**Figure 8:** Average MAE comparison by primary use

Figure 9: Average TT comparison by primary use

smart grid systems' control systems work optimally in real time and computational resources are limited, FFNN models may be preferable. The selection of two fixed architectures and completion of a total of 192 predictions proved that both FFNN and LSTM are flexible and scalable. In a real-world business case the focus must be on minimizing the MAE (Figure 9).

6. CONCLUSION

The increased predictability of problems related to the production, distribution, and consumption of electrical energy offers a good foundation for increasing the adoption of energy obtained from renewable sources, economical optimization by introducing flexible pricing policies, and reducing electricity consumption. At the same time, Six Sigma DMAIC methodology ensures that the initial setup for the problem statement and needs, goals, and possible blocking points in a clear and simple yet powerful framework. Modeling real-world data using ANNs under a Six Sigma DMAIC cycle—a robust data mining framework—proved to be successful in terms of performance and training speed. Moreover, the flexibility and scalability have been proven by maintaining the level of performance for extreme scenarios in which the available data consisted of either thousands or millions of records.

Increasing the performance of the models it is a matter of having more data with better quality. Also, a finer discretization of the primary use will bring together use cases likely to have the same patters. Comparing the results by the primary use can be highlighted the fact that one model can capture the phenomena better than the other one, with the mention that on a bigger resolution (e.g., prediction by primary use and building) both models may have better and closer performance. Although one solution might look initially the best before deploying it into production a rigorous validation process must be conducted.

By selecting R^2 as one of the metrics comparisons with other similar research papers can be made and a benchmark may be set. Moreover, by listing all models' parameters, software and hardware configuration will allow other researchers to perform the same experiments.

Comparing the results with those obtained by other researchers was not possible due to the way the metrics are typically selected; specifically, the metrics allow a comparison of models that are trained using the same data set, but do not allow a comparison

of models trained on different data sets. In this regard, the use of the coefficient of determination, the complete description of the models' parameters, and the software and hardware configuration will allow other researchers to use this article for comparative studies. The performance of the models might be increased by adding exogenous variables, such as wind speed, wind shear, ambient temperature and pressure, dew point temperature and humidity.

The freshly approved Romanian Recovery and resilience plan provides 41% of the total amount for green transition and 21% for digital transition. In the key of this achievement researchers will be able to continue their work.

REFERENCES

Abdeljaber, O., Avci, O., Kiranyaz, S., Boashash, B., Sodano, H. and Inman, D. (2017), 1-D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data. Neurocomputing, 275, 1308-1317.

Al-Ali, A.R. (2016), Internet of things role in the renewable energy resources. Energy Procedia, 100, 34-38.

Ashton, K. (2009), That "internet of things" thing: In the real world things matter more than ideas. RFID Journal, 22(7), 97-114.

Bengio, Y., Simard, P., Frasconi, P. (1994), Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157-166.

Botea, R. (2020), Energiile Regenerabile au Acoperit 42% din Consumul de Energie al României, cu 10 Puncte Procentuale Peste Media Europeană. Available from: https://www.zf.ro/eveniment/energiile-regenerabile-au-acoperit-42-din-consumul-de-energie-al-romaniei-cu-10-puncte-procentuale-peste-media-europeana-18764797 [Last accessed on 2020 Oct 19].

Chen, L., Lai, X. (2011), Comparison between ARIMA and ANN Models Used in Short-term Wind Speed Forecasting. In: IEEE, 2011 Asia-Pacific Power and Energy Engineering Conference. Wuhan, China, 25-28 March 2011.

Cho, K., van Merriënboer, B., Gulcehre, C., Bougares, F., Schwenk, H., Bengio, Y. (2014), Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar, 25-29 October 2014.

Chollet F. (2017), Deep Learning with Python. Greenwich, CT: Manning Publications.

Ding, M., Zhou, H., Xie, H., Wu, M., Nakanishi, Y., Yokoyama, R. (2019), A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. Neurocomputing, 365, 54-61.

Eldali, F., Hansen, T., Suryanarayanan, S., Chong, E. (2016), Employing ARIMA models to improve wind power forecasts: A case study in ERCOT. In: IEEE, 2016 North American Power Symposium (NAPS). Denver, CO, 18-20 September 2016.

End to End Machine Learning School. (2020), Convolution in One Dimension for Neural Networks. Available from: https://e2eml.school/convolution_one_d.html [Last accessed on 2020 Dec 20].

European Commission. (2020), EU Climate Policies and the European Green Pact. Available from: https://ec.europa.eu/clima/policies/eu-climate-action_ro [Last accessed on 2020 Oct 20].

European Court of Auditors. (2019), Wind and Solar Energy for Electricity Generation: Significant Action is Needed to Achieve EU Targets. Available from: https://op.europa.eu/webpub/eca/special-reports/wind-solar-power-generation-8-2019/ro/index.html#h2table5 [Last

accessed on 2020 Dec 19].

Fawcett, T., Provost, F. (2013), Data Science for Business. Newton, MA: O'Relly.

Fukuoka, R., Suzuki, H., Kitajima, T., Kuwahara, A., Yasuno, T. (2018), Wind speed prediction model using LSTM and 1D-CNN. Journal of Signal Processing, 22(4), 207-210.

Hanski, J., Uusitalo, T., Vainio, H., Kunttu, S., Valkokari, P., Kortelainen, H., Koskinen, K. (2018), Smart Asset Management as a Service Deliverable 2.0. Available from: http://doi.org/10.13140/RG.2.2.31027.94244 [Last accessed on 2020 Oct 19].

Hochreiter, S., Schmidhuber, J. (1997), Long short-term memory. Neural Computation, 9(8), 1735-1780.

IBM. (2019), IBM SPSS Modeler CRISP-DM Guide: CRISP-DM Help Overview. Available from: https://www.ibm.com/support/knowledgecenter/SS3RA7_sub/modeler_crispdm_ddita/clementine/crisp_help/crisp_overview.html [Last accessed on 2020 Oct 19].

Kingma, D., Ba, J. (2015), Adam: A Method for Stochastic Optimization. In: 3rd International Conference for Learning Representations, San Diego, CA, 7-9 May 2015.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J. (2020), 1D Convolutional Neural Networks and Applications: A Survey. Mechanical Systems and Signal Processing, 151, 107398.

Krishna, P.G., Ravi, K.S., Kishore, K.H., Veni, K.K., Rao, K.N.S., Prasad, R.D. (2018), Design and development of bi-directional IoT gateway using ZigBee and Wi-Fi technologies with MQTT protocol. International Journal of Engineering and Technology, 7(28), 125-129.

Le Cun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D. (1990), Handwritten Digit Recognition with a Back-Propagation Network. In: Touretzky, D. editor. Advances in Neural Information Processing Systems (NIPS 1989), Denver, CO, 27-30 November 1989. Burlington: Morgan Kaufmann.

Liu, Y., Ding, S., Jia, W. (2020), A novel prediction method of complex univariate time series based on k-means clustering. Soft Computing, 24, 16425-16437.

McCulloch, W.S., Pitts, W. (1943), A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115-133.

Neon Neue Energieökonomik. (2020), Open Power Systems Data: Load, Wind and Solar, Prices in Hourly Resolution. Available from: https://data.open-power-system-data.org/time_series/2020-10-06 [Last accessed on 2020 Oct 19].

Pant, P., Garg, A. (2016), Forecasting of short term wind power using ARIMA method. International Journal for Research in Applied Science and Engineering Technology, 4(3), 356-361.

Pascanu, R., Gulcehre, C., Cho, K., Bengio, Y. (2014), How to construct deep recurrent neural networks. In: ICLR, 2nd International Conference on Learning Representations. Banff, Canada, 14-16 April 2014.

Rumelhart, D., Hinton, G.E., Williams, R.J. (1986), Learning representations by back-propagating errors. Nature, 323, 533-536.

Sava, J.A. (2020), Onshore Wind Energy Capacity in Romania 2008-2019. Available from: https://www.statista.com/statistics/870766/onshore-wind-energy-capacity-in-romania [Last accessed on 2020 Dec 21].

Singh, V. (2020), 10 Benefits of Using Cloud Storage. Available from: https://cloudacademy.com/blog/10-benefits-of-using-cloud-storage [Last accessed on 2020 Oct 19].

Wang, J., Hu, J. (2015), A robust combination approach for short-term wind speed forecasting and analysis combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. Energy, 93, 41-56.

Yun, M., Yuxin, B. (2010), Research on the architecture and key technology of Internet of Things (IoT) applied on the smart grid. In: IEEE, 2010 International Conference on Advances in Energy Engineering, Beijing, China, 19-20 June 2010.

Zhang, A., Lipton Z.C., Li, M., Smola, A.J. (2020), Dive into Deep Learning. Available from: https://d2l.ai/chapter_convolutional-neural-networks/index.html [Last accessed on 2020 Dec 20].