Gas Consumption Forecasting Using Machine Learning Methods and Taking Into Account Climatic Indicators

Grigorii Shymchuk[a], Iaroslav Lytvynenko[a], Roman Hromyak[a], Sergii Lytvynenko[a] and Volodymyr Hotovych[a]

[a] *Ternopil Ivan Puluj National Technical University, 56, Ruska Street, Ternopil, 46001, Ukraine*

Abstract

This paper examines a method for forecasting gas consumption based on the application of an additive mathematical model. One of the components of the mathematical model is a cyclic random process. The forecast is constructed using machine learning methods, specifically gradient boosting and decision trees. When constructing the forecast, climatic factors that influence natural gas consumption were taken into account. It has been established that the developed forecasting method, which considers the cyclic component of gas consumption in the model, exhibits higher accuracy compared to the known method.

Keywords

Cyclic process, gas consumption process, statistical processing, cyclic random process, forecasting, machine learning methods

1. Introduction

The increasing demand for energy consumption in various industries creates the prerequisites for the development of numerous means and methods of optimization and control over the use of these resources. Natural gas consumption is no exception. The availability of a significant amount of data on gas consumption allows for the efficient application of machine learning and decision-making methods to improve gas consumption processing and forecasting tasks. Moreover, decision-making methods are also used to assist operators in making informed decisions regarding the optimization of gas consumption processes. These methods prove their effectiveness in monitoring gas consumption processes, reducing electricity costs, and minimizing environmental impact. In this context, the subject of processing and forecasting natural gas consumption attracts significant interest from researchers and practitioners who are interested in developing innovative solutions to improve results. New mathematical models and gas consumption processing and forecasting methods using machine learning techniques are of particular interest.

2. Analysis of recent research

Many researchers are engaged in the development of gas consumption forecasting methods. For instance, in [1], three different types of models for short-term natural gas consumption forecasting on a daily basis were developed. The authors used a sigmoid regression model, a feedforward neural network model, and a recurrent neural network model. The results obtained indicate that the accuracy of the forecast based on artificial neural networks was higher than the accuracy of the forecast based on the sigmoid regression model.

In [2], mathematical models for gas consumption forecasting were proposed, including models based on logistic curves, statistical models, econometric models, neural networks, and genetic algorithms.

CITI'2023: 1st International Workshop on Computer Information Technologies in Industry 4.0, June 14–16, 2023, Ternopil, Ukraine
EMAIL: gorych@gmail.com (G. Shymchuk); iaroslav.lytvynenko@gmail.com (I. Lytvynenko); hromjakr@gmail.com (R. Hromyak); serge.lytvynenko@gmail.com (S. Lytvynenko); gotovych@gmail.com (V. Hotovych)
ORCID: 0000-0003-2362-7386 (G. Shymchuk), 0000-0001-7311-4103 (I. Lytvynenko); 0000-0002-3368-2613 (R. Hromyak); 0009-0003-0868-9156 (S. Lytvynenko);0000-0003-2143-6818 (V. Hotovych)

Machine learning methods such as artificial neural networks (ANN) and support vector regression (SVR) were used to forecast natural gas consumption based on secondary data collected under various influencing factors.

In [3], gas consumption forecasting was proposed using hybrid neural networks with the approach of artificial bee colonies.

The study presented in [4] provided gas consumption forecasting results obtained using artificial neural networks (ANN). Specifically, the MLP model (multilayer perceptron model) was used, and the training was performed using data describing actual gas consumption.

In [5], four different forecasting models using a block structure and combining its elements to create three hybrid models were developed. Overall, the application of models based on artificial neural networks showed higher accuracy compared to gradient boosting models, and the combined (hybrid) models demonstrated significantly better results in forecasting.

The research conducted in [6] aimed to develop efficient neural network models for predicting natural gas energy demand by utilizing various quantitative and qualitative input data, including social factors. The neural network model had a deep, complex architecture with multiple layers and nodes. Three different approaches were tested, including artificial neural networks (ANN), long short-term memory (LSTM) models, and a proposed implementation of deep neural networks (DNN). The proposed DNN implementation outperformed the ANN and LSTM approaches.

In [7], short-term forecasting for the next day was investigated. The authors proposed a model that employed wavelet transformation, genetic algorithms, and neural network methods for demand forecasting.

In [8], neural networks and a sliding window approach were used for forecasting gas consumption for the next day, and it was found that although the algorithm training was demanding, the applied algorithm was relatively simple and did not require extensive computational resources.

The article in [9] focused on the importance of gas consumption forecasting in the residential sector, comparing four different models, two neural network models, and two econometric models to find the most accurate method for forecasting monthly natural gas demand. The selected forecasting period spanned 24 months, and the comparison aimed to assess the strength of different models in each specific problem.

In [10], the use of deep neural networks for short-term gas consumption forecasting was proposed, comparing them with a linear regression model and a traditional artificial neural network model.

3. Main part

In [11], an additive mathematical model of the gas consumption process was proposed, one of the components of which was a cyclic random process. Let us briefly consider this model. The model of the cyclic random process of gas consumption $\xi'(\omega,t)$ is presented in the form of an additive model (1) consisting of three components:

$$\xi'(\omega,t) = \xi(\omega,t) + f_{tr}(t) + f_{rem}(\omega'',t),\ t \in \mathbf{W}, \omega \in \mathbf{\Omega}, \omega'' \in \mathbf{\Omega}'', \tag{1}$$

where $\xi(\omega,t)$ – cyclic component, $f_{tr}(t)$ - trend function, $f_{rem}(\omega'',t)$ - stochastic function remainder.

In practice, we have to deal with discrete data, so it is convenient to work with the mathematical model (1) in this form:

$$\xi'_{\omega}(l) = \xi_{\omega}(l) + f_{tr}(l) + f_{rem\omega''}(l),\ l \in \mathbf{W} = \mathbf{D}, \tag{2}$$

where $\xi_{\omega}(l)$ – implementation of the cyclical component of the gas consumption process, $f_{tr}(l)$ - trend function, $f_{rem\omega''}(l)$ - function stochastic remainder, l - discrete readout of the gas consumption process.

Also, work [11] shows how each of the components of the additive mathematical model is obtained, so we will consider $\xi_{\omega}(l)$ - cyclic component of the mathematical model (2), which carries information about the gas consumption process, we will present it in the form of:

$$\xi_{\omega}(l) = \sum_{i=1}^{C} f_i(l),\ l \in \mathbf{W}, \tag{3}$$

where C – the number of segments-cycles of the cyclic process of gas consumption. $\mathbf{W}$ – region of definition of the cyclic process of gas consumption, and the region of its values, for the case of the stochastic approach, is the Hilbert space of random variables set on one probability space $\left(\xi_{\omega}(l) \in \mathbf{\Psi} = \mathbf{L}_2(\mathbf{\Omega}, \mathbf{P})\right)$. In construction (3), there are segments-cycles $f_i(l)$ of the cyclic process of gas consumption are determined through indicator functions, i.e

$$f_i(l) = \xi_{\omega}(l) \cdot I_{\mathbf{W}_i}(l), i = \overline{1, C},\ l \in \mathbf{W}. \tag{4}$$

At the same time, the indicator functions that distinguish segments-cycles were defined as:

$$I_{\mathbf{W}_i}(l) = \begin{cases} 1, l \in \mathbf{W}_i, \\ 0, l \notin \mathbf{W}_i. \end{cases},\ i = \overline{1, C}, \tag{5}$$

where $\mathbf{W}_i$ – the domain of the indicator function, which in the case of a discrete signal, i.e $\mathbf{W} = \mathbf{D}$, is equal to a discrete plural of counts

$$\mathrm{W}_i = \left\{ l_{i,j},\ j = \overline{1, J} \right\},\ i = \overline{1, C}. \tag{6}$$

The segmental cyclic structure $\hat{\mathbf{D}}_c$ is taken into account by a set of time counts $\{l_i\}$ or $\{l_{i,j}\}$, $i = \overline{1, C}$ $j = \overline{1, J}$, where J – the number of discrete counts per cycle. This form of the mathematical model (6) takes into account the rhythm of the cyclic process of gas consumption through a continuous function of the rhythm $T(l, n)$, namely:

$$I_{\mathbf{W}_i}(l) = I_{\mathbf{W}_{i+n}}\left(l + T(l, n)\right),\ i = \overline{1, C}, n = 1, l \in \mathbf{W}. \tag{7}$$

To assess the rhythm function $T(l, n)$ define the segmental structure of the gas consumption process [12], as $\hat{\mathbf{D}}_c = \{l_i, i = \overline{1, C}\}$, after that, the rhythmic structure is evaluated and statistical processing methods are applied in the further processing steps.

In work [11] it is shown that for an adequate description of the real process of gas consumption, it is necessary to take into account changes in the amplitude of the load on segments-cycles, which are caused by various climatic factors such as: temperature, pressure, wind force and others.

Let's denote these factors by a vector $\Xi_N = \left\{ \eta_{1\mathbf{W}_i}(l), \eta_{2\mathbf{W}_i}(l), \ldots \eta_{N\mathbf{W}_i}(l), l \in \mathbf{W}_i, i = \overline{1, C} \right\}$, and in model (4) we take into account its multiplicative component, taking into account indicator functions and scale coefficients of gas consumption amplitude, i.e., the component of mathematical model (4) will have the form:

$$f_i(l) = \xi_{\omega}(l) \cdot \alpha_{\mathbf{W}_i}(l) \cdot I_{\mathbf{W}_i}(l), i = \overline{1, C},\ l \in \mathbf{W}. \tag{8}$$

In formula (8) of the component $\alpha_{\mathbf{W}_i}(l)$, reflects the scale coefficients of the amplitude of gas consumption on each segment-cycle of the cyclic process, namely:

$$\alpha_{\mathbf{W}_i}(l) = \begin{cases} \alpha_i, l \in \mathbf{W}_i, \\ 0, l \notin \mathbf{W}_i. \end{cases},\ i = \overline{1, C}, \tag{9}$$

where α_i – scale coefficients of gas consumption amplitude on each i-th segment-cycle are defined as follows:

$$\alpha_i = \frac{\alpha_{i\max}}{\alpha_{aver}},\ i = \overline{1, C}, \tag{10}$$

where $\alpha_{i\max}$ – the maximum value of the amplitude of gas consumption on i-th segment-cycle (determined at the stage of segmentation of the cyclic process of gas consumption), α_{aver} - the average value of the amplitude of gas consumption (the maximum value of the amplitude of the estimate of mathematical expectation, determined at the stage of statistical processing of the cyclic process of gas consumption).

For the climate indicator vector $\Xi_N = \left\{ \eta_{n\mathbf{W}_i}(l), n = \overline{1, N}, l \in \mathbf{W}_i, i = \overline{1, C} \right\}$ a fair ratio that takes into account the climatic parameters that affect gas consumption in each segment-cycle of the cyclic process, namely:

$$\eta_{n\mathbf{W}_i}(l)=\begin{cases}\eta_{ni},\ l\in\mathbf{W}_i,\\ 0,\ l\notin\mathbf{W}_i.\end{cases},\ i=\overline{1,C},\ n=\overline{1,N}, \tag{11}$$

where η_{ni} climatic parameter (value) which corresponds n –th climatic indicator (temperature, pressure, etc.) on each i - th segment-cycle of the gas consumption process.

Let's consider the developed method of forecasting gas consumption taking into account climate indicators using machine learning methods. Figure 1 shows the input data of gas consumption during 2006-2021.

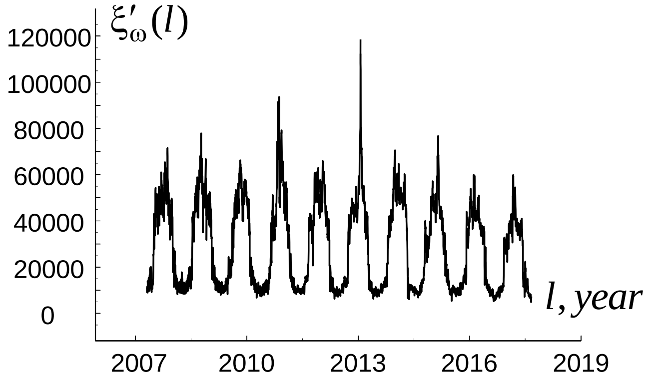


Figure 1: Fragment of input implementation of cyclic gas consumption process

Figure 2 shows two components of the mathematical model: the trend component and the stochastic residual. The cyclic component is presented in Figure 3, and it will be used specifically for forecasting purposes.

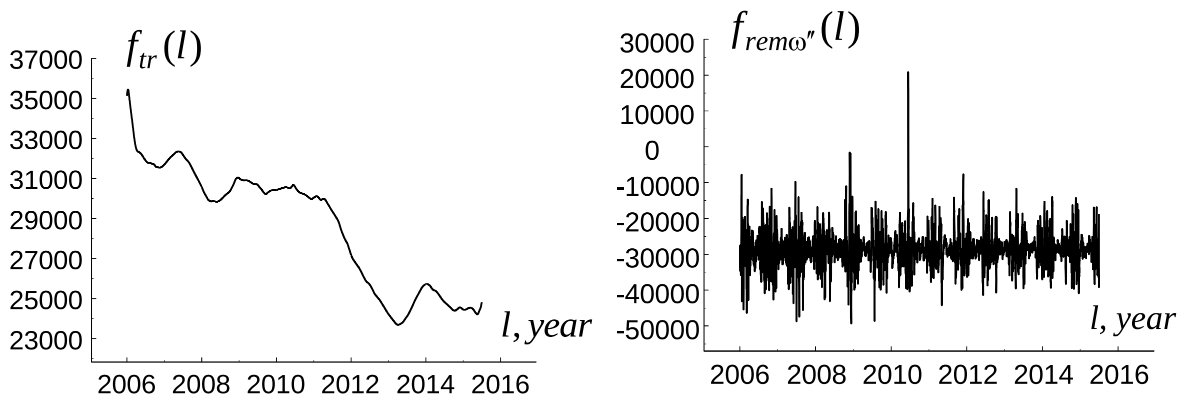


Figure 2: Estimated components of the mathematical model: a) trend component; b) stochastic residue

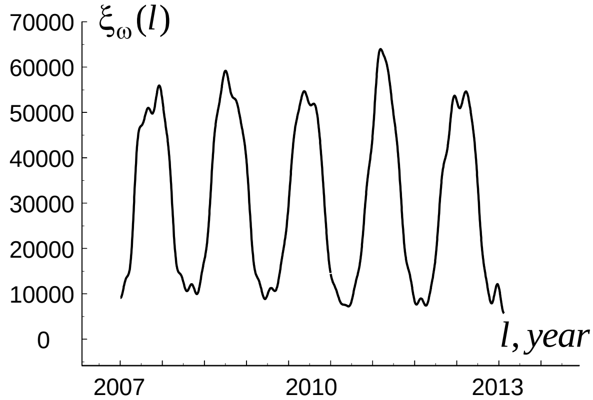


Figure 3:. Fragment of the studied implementation of the gas consumption process for the case of segmentation by cavity

Examples of climatic indicators that affect the gas consumption process are shown in Figure 4, since climatic indicators are recorded at different intervals, the figure shows only the maximum values of climatic indicators that affect the gas consumption process.

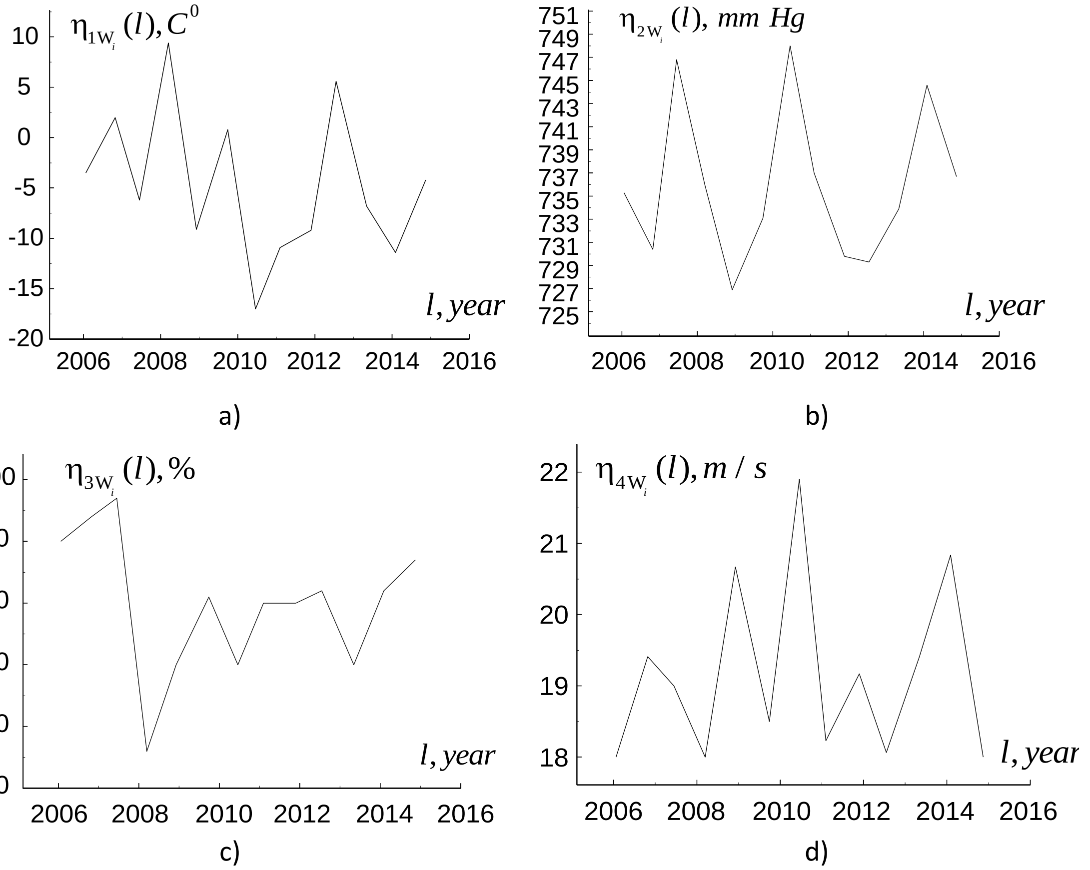
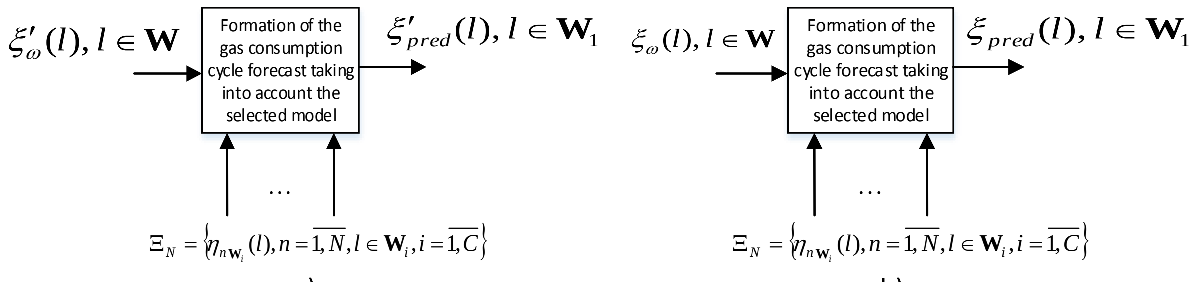


Figure 4: Examples of four climate indicators: a) air temperature (at a height of 2 meters above the ground); b) atmospheric pressure at the level of the station; c) relative humidity (at a height of 2 meters above the ground); d) the maximum value of the wind gust (at a height of 10-12 meters above the earth's surface during the 10-minute period immediately preceding the observation period)

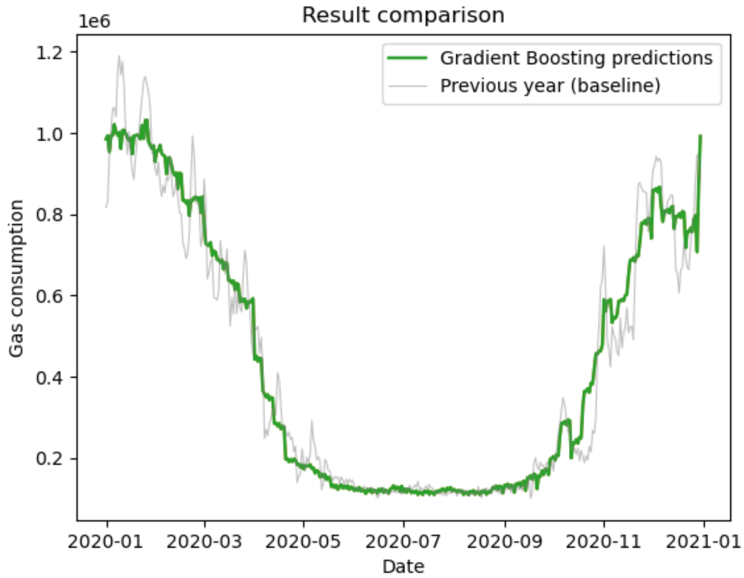
Figure 5 shows in the form of building blocks two approaches to forecast gas consumption values using different input data for training.



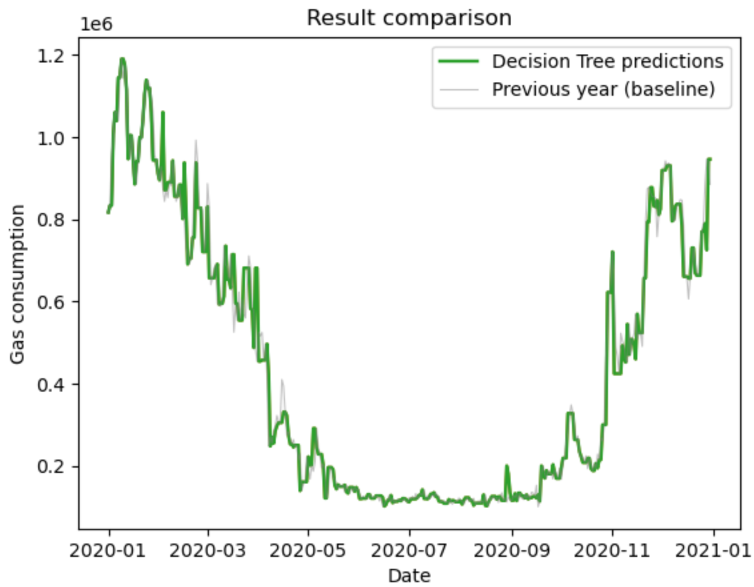
a) b)

Figure 5: Structural diagram of formation of gas consumption forecasting using different training data: a) incoming data in the form of the gas consumption process; b) incoming data for training in the form of the cyclic component of gas consumption

We will apply two models and two sets of data to forecast gas consumption values and compare the obtained forecast results taking into account climatic indicators. To this end, for the first approach, we will take a neural network model of gradient boosting and a model based on decision trees for training, and input data in the form of the gas consumption process presented in Figure 1 (without processing). At the same time, we will get the forecast results shown in Figure 6. When forecasting, the root mean square error for the case of applying a forecast based on gradient boosting is 0.8915, and for the case of a forecast based on decision trees is 0.937.



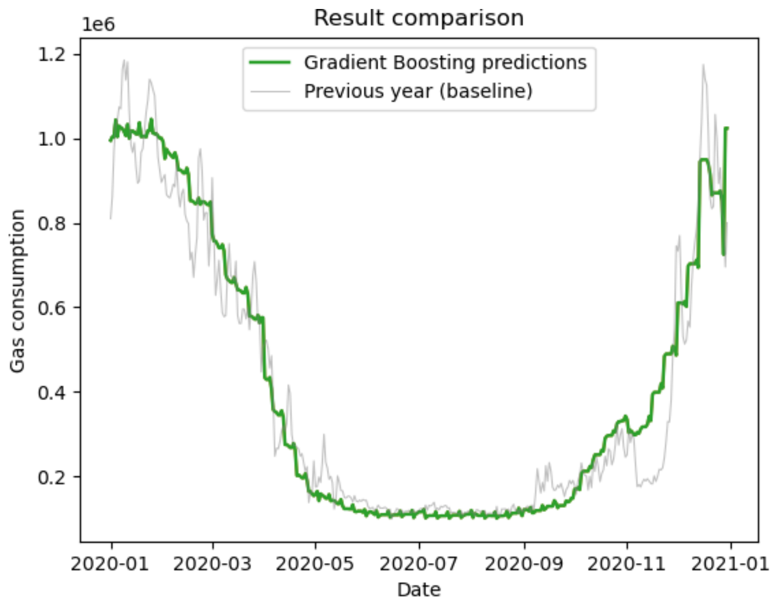
a)



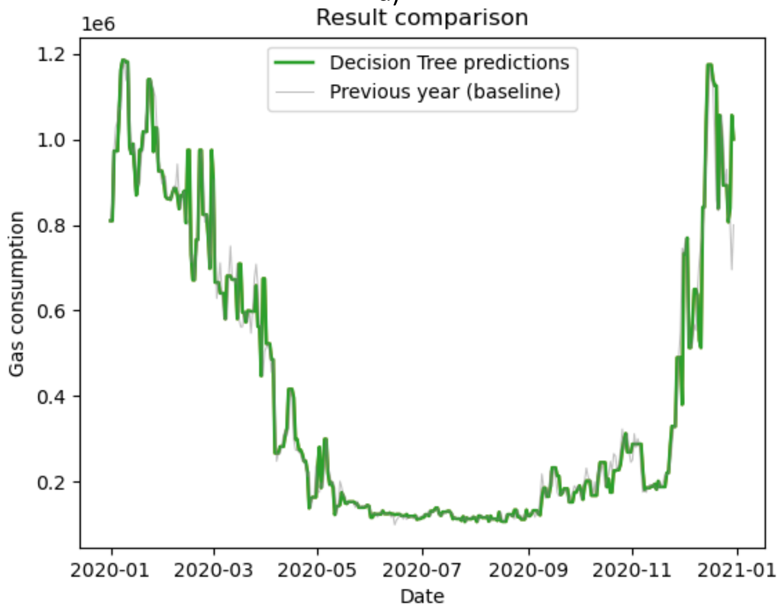
b)

Figure 6: Forecasted gas consumption values and gas consumption input data (learning from gas consumption input data, Figure 1): a) prediction based on the application of the gradient boosting model; b) prediction based on the application of the decision tree model

For the second approach, we will use the neural network model of gradient boosting and the model based on decision trees for training, and the input data in the form of the gas consumption process presented in Figure 3 (processed cyclic component). At the same time, we will get the forecast results shown in Figure 7. When forecasting, the root mean square error for the case of applying a forecast based on gradient boosting is 0.904, and for the case of a forecast based on decision trees, it is 0.953.



a)



b)

Figure 7: Forecasted gas consumption values and input gas consumption data (learning based on the cyclic component of gas consumption, Figure 3): a) prediction based on the application of the gradient boosting model; b) prediction based on the application of the decision tree model

Let's proceed to the discussion of the obtained results.

4. Discussion of obtained results

From the results obtained, it follows that the root mean square error for the case of applying the gradient boosting model is larger than in the case of applying the model based on decision trees. In addition, training the network based on the input data in the form of a cyclic component of the gas consumption process and forecasting on its basis showed that the root mean square error is significantly lower than the others, which indicates the accuracy of forecasting and makes the developed forecasting method better than others.

5. Conclusions

In this paper, a gas consumption forecasting method was developed based on the application of gradient boosting machine learning methods and a method based on decision trees. From the obtained results, it was found that the forecasting method based on the model of decision-making trees more accurately forms the forecast of gas consumption values. In addition, we found that the forecast based on the learning of the input data in the form of a cyclic component of gas load gives better results compared to the forecast based on the input data of gas consumption presented in Figure 1. Because the developed method allows for a more accurate forecast in contrast to the use of other models.

It is planned to conduct a study on the formation of a forecast based on the application of a model in the form of a random forest and to compare the obtained results with the developed forecast method. It is planned to develop an appropriate software complex for predicting gas consumption too.

6. References

[1] J. Ravnik, J. Jovanovac, A. Trupej, N. Vistica, M. Hribersek. A sigmoid regression and artificial neural network models for day-ahead natural gas usage forecasting. / Ravnik J., Jovanovac J., Trupej A., Vistica N., Hribersek M. / Cleaner and Responsible Consumption, Volume 3, 2021, 100040, ISSN 2666-7843, ttps://doi.org/10.1016/j.clrc.2021.100040.

[2] Sanjeet Singh, Pooja Bansal, Mosharrof Hosen, Sanjeev K. Bansal. Forecasting annual natural gas consumption in USA / Singh Sanjeet, Bansal Pooja, Hosen Mosharrof, Bansal Sanjeev K. / Application of machine learning techniques- ANN and SVM. Resources Policy, Volume 80, 2023, 103159, ISSN 0301-4207, https://doi.org/10.1016/j.resourpol.2022.103159.

[3] Akpinar Mustafa, M. Fatih Adak, Yumusak Nejat. 2016. Forecasting natural gas consumption with hybrid neural networks artificial bee colony. In: 2016 2nd International Conference on Intelligent Energy and Power Systems. IEPS, pp. 1–6.

[4] Jolanta Szoplik. Forecasting of natural gas consumption with artificial neural networks. Energy, Volume 85, 2015, Pages 208-220, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2015.03.084.

[5] Vinayak Sharma, Umit Cali, Bhav Sardana, Murat Kuzlu, Dishant Banga, Manisa Pipattanasomporn. Data-driven short-term natural gas demand forecasting with machine learning techniques. Journal of Petroleum Science and Engineering,Volume 206, 2021, 108979, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2021.108979.

[6] Anagnostis, Athanasios & Papageorgiou, Elpiniki & Bochtis, Dionysis. (2020). Application of Artificial Neural Networks for Natural Gas Consumption Forecasting. Sustainability. 12. 6409. 10.3390/su12166409.

[7] Ioannis P. Panapakidis, Athanasios S. Dagoumas. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. Energy, 118 (2017), pp. 231-245.

[8] Mustafa Akpinar, M. Fatih Adak, Nejat Yumusak. Day-ahead natural gas demand forecasting using optimized ABC-based neural network with sliding window technique: the case study of regional basis in Turkey. Energies, 10 (6) (2017).

[9] Mohamad Hossein Safiyari, Saeed Shavvalpour, Sina Tarighi. From traditional to modern methods: Comparing and introducing the most powerful model for forecasting the residential natural gas demand. Energy Reports, Volume 8, 2022, Pages 14699-14715, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2022.10.397.

[10] Merkel, Gregory & Povinelli, Richard & Brown, Ronald. (2018). Short-Term Load Forecasting of Natural Gas with Deep Neural Network Regression †. Energies. 11. 2008. 10.3390/en11082008.

[11] Simulation of gas consumption process based on the mathematical model in the form of cyclic random process considering the scale factors Lytvynenko, I., Lupenko, S., Kunanets, N., Shymchuk, G., Hotovych, V. CEUR Workshop Proceeding, 2021, 3039, pp. 97–106.

[12] Additive mathematical model of gas consumption process / Iaroslav Lytvynenko, Serhii Lupenko, Oleh Nazarevych, Hryhorii Shymchuk, Volodymyr Hotovych // Scientific Journal of TNTU. Tern.: TNTU, 2021. Vol 104. No 4. pp. 87–97.