Mediterranean Smart Cities Conference (MSCC 2024)
2-4 May, 2024, Martil, Morocco

Enhanced Cascade Schemes for Advancing Machine Learning-Based Prediction of Heating and Cooling Loads in Residential Buildings

Ivan Izonin
Department of Artificial Intelligence / Department of Civil Engineering
Lviv Polytechnic National University
University of Birmingham
Lviv, Ukraine / Birmingham, United Kingdom
ivanizonin@gmail.com
0000-0002-9761-0096

Roman Tkachenko
Department of Publishing Information Technologies
Lviv Polytechnic National University
Lviv, Ukraine
roman.tkachenko@gmail.com
0000-0002-9802-6799

Ostap Shcherbii
Department of Artificial Intelligence
Lviv Polytechnic National University
Lviv, Ukraine
ostap.shcherbii.knm.2020@lpnu.ua
0009-0007-4029-2932

Roman Muzyka
Department of Artificial Intelligence
Lviv Polytechnic National University
Lviv, Ukraine
roman.muzyka.mknssh.2022@lpnu.ua
0009-0001-3884-8280

Asaad Faramarzi
Department of Civil Engineering
University of Birmingham
Birmingham, United Kingdom
a.faramarzi@bham.ac.uk
0000-0001-6964-5577

Stergios Aristoteles Mitoulis
Department of Civil Engineering
University of Birmingham
Birmingham, United Kingdom
s.a.mitoulis@bham.ac.uk
0000-0001-7201-2703

Abstract—**The contemporary imperative of addressing the challenge of energy-efficient building design underscores the significance of employing advanced methodologies. Leveraging machine learning (ML) tools has emerged as a viable strategy to streamline the modeling phase, a pivotal step in the design process for such structures. Despite the potential benefits, the precision of ML methods may be compromised due to various factors. This research introduces two modifications to the cascade correction schemes applied to regression models, aiming to enhance the accuracy of the selected ML method. These enhancements are grounded in the concept of establishing a cascade of two ML algorithms utilizing rational fractions (direct approach). The modification involves an additional modeling step that focuses on elucidating the relationships among independent inputs for the second ML method. This is achieved through nonlinear expansion using the Wiener polynomial. Such an approach contributes to an augmented accuracy of the ML algorithm, thereby elevating the overall precision of the cascade. The modeling of these modifications was conducted using a real-world dataset, specifically addressing the prediction of heating and cooling loads in residential buildings. Optimal parameters for the proposed modifications were identified, showcasing a noteworthy enhancement in accuracy compared to existing methods.**

Keywords—cascade scheme; resilient AI; rational fractions; Ito decomposition; building performance [1]

I. Introduction

The heating and cooling load prediction task in residential buildings is crucial in modern conditions, given the general paradigm shift towards sustainable development and energy efficiency in construction [1]. The increasing awareness of buildings' environmental impact and the continuous rise in energy resource costs emphasize the necessity to develop advanced technologies for optimizing energy consumption in residential premises [2], [3].

The effectiveness of solving this task in residential buildings becomes a key aspect of managing a comfortable climate and rational use of energy [4], [5]. The development of accurate and reliable prediction methods [6] is critically important for enhancing energy efficiency, reducing emissions, and providing comfortable living conditions for building residents. Therefore, research in this direction is exceptionally significant in the context of implementing innovative approaches and technologies aimed at creating efficient heating and cooling systems that adhere to high standards of energy efficiency and environmental responsibility.

The utilization of machine learning in predicting heating and cooling loads presents opportunities for improving the efficiency of designing and operating

[1] This work is funded by the European Union's Horizon Europe research and innovation program under grant agreement No 101138678, project ZEBAI (Innovative methodologies for the design of Zero-Emission and cost-effective Buildings enhanced by Artificial Intelligence)

comfortable climate systems in buildings [7]. However, the accuracy of machine learning methods relies on various factors, notably the quantity and quality of input data, the appropriate selection of the actual algorithm for approximation, and the correct configuration of its parameters [8]. Nevertheless, in the operational mode of the method, data with outliers and anomalies, data drift, and distorted values of specific parameters may be encountered as a response to the external environment [9], [10]. In such instances, the machine learning method may exhibit a notable decrease in the predictive accuracy of results [11], [12]. Consequently, the necessity arises to minimize the impact of such situations and alleviate the resulting challenges.

This paper aims to enhance the accuracy of machine learning methods in addressing the prediction task of heating and cooling loads in residential buildings.

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

- We modified two two-step cascade regression model parameters refinement schemes (direct approach) due to the use of nonlinear expansion of inputs for the second machine learning method in the cascade scheme using the Wiener polynomial;
- We identified optimal operating parameters for both modifications, demonstrating an enhanced operational efficiency compared to the baseline cascade schemes in addressing the prediction task of heating and cooling loads in residential buildings.

II. Materials and Methods

A. Modification 1

Let the given tabular dataset be approximated by a certain machine learning algorithm:

$$F(x_{i,1},\ldots,x_{i,m}) \to y_i \tag{1}$$

The accuracy of his predictions y_i^{pred}, however, is not satisfactory. The task of improving the accuracy, as discussed in [13] and [14], is addressed through the construction of two-level cascading schemes for refining the prediction results. This is achieved by employing a rational fraction, as described in [13]:

$$y_i = \frac{y_i^{pred}}{1+F(x_{i,1},\ldots,x_{i,n})} \tag{2}$$

Based on (2), we can write the following expression:

$$F(x_{i,1},\ldots,x_{i,n}) = z_i,\ where\ z_i = \frac{y_i^{pred}}{y_i} - 1, \tag{3}$$

The accuracy correction of the predicted results (1) can be achieved through the application of a second machine learning algorithm to approximate (3):

$$g(x_{i1},\ldots,x_{in}) \to z_i, \tag{4}$$

However, considering that the value of z_i, depends on the accuracy of work (1) and may be quite small, this paper proposes a nonlinear transformation of the inputs $x_{i,1},\ldots,x_{i,n}$ using a second-degree Wiener polynomial:

$$Y(x_1,\ldots,x_n) = a_i + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=i}^{n} a_{i,j} x_i x_j \tag{5}$$

This additional modeling captures relationships among the set's independent variables, expanding the input data space. Leveraging the strong approximation properties of (5), enhances the accuracy of the results for (4).

By using the predicted outcomes from (1), performing (5), and obtaining the results from (4), it is possible to obtain the overall result of the modified method:

$$y_i^{(1)} = y_i^{pred} / (z_i^{pred} + 1) \tag{6}$$

The structural and functional scheme of this approach is shown in Fig. 1

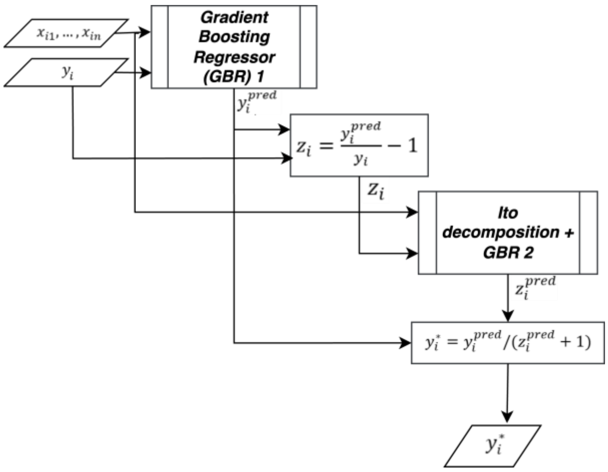

Figure 1. Flowchart of the basic cascade scheme modification

This is the first modification of the cascade scheme from[13] proposed in this paper.

B. Modification 2

The paper [14] introduced an enhanced implementation algorithm for the approach presented in [13], incorporating an additional procedure to rectify prediction results. This happens because the result of operation (6) can be iteratively substituted into (3) instead y_i^{pred} until the specified stop criterion is met. We adopted this approach in our research, implementing it precisely according to the modification 1 proposed in this paper.

The structural and functional scheme of operation of modification 2 is illustrated in Fig. 2.

Similar to the first case, modification 2 relies on the use of nonlinear expansion of the inputs using (5). The combination of this with the use of the iterative procedure from Fig. 2 is expected to enhance the accuracy of modification 2.

The stopping criterion in this case is determined by the iteration at which the error selected by the user in the application mode of the method starts to increase.

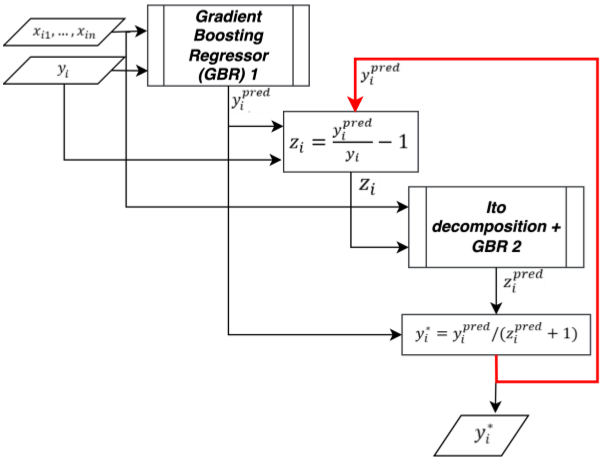

Figure 2. Flowchart of the iterative cascade scheme modification

III. Modeling and Results

In [15], a detailed investigation was conducted on the effectiveness of four machine learning methods for addressing the prediction task of heating and cooling loads in residential buildings. The Gradient Boosting Regressor emerged as the most effective method. In this paper, the simulation of the operation of both modified cascade schemes was carried out using this method at the initial level of the cascade. The primary objective of this paper is to enhance the accuracy of the proposed approach.

A. Datasets descriptions

The experimental study was carried out using a real-world dataset from [15] through our software. The author in [15] modeled the energy efficiency of the building (heating and cooling loads) using machine learning methods. He uses a fairly short dataset (768 vectors) characterized by 6 independent variables and two dependent variables that should be predicted. After cleaning and preparation, the dataset was randomly split into training and test sets in the ratio of 80% to 20%. Normalization of the dataset was performed using MinMaxScaler for our research.

B. Parameters of the investigated methods

To preserve the transparency of the experiment, we used the same parameters of the Gradient Boosting regressor that were used in [15]. The basic cascade schemes were created to increase the accuracy of the latter. The identical regressor with consistent parameters was chosen as the machine learning model used at the second level of the cascade (learning_rate=0.1, n=250, depth=5, samples split=2, leaf=3, subsample=1.0). This approach ensures the possibility of accurate reproducibility of our experiments.

The first modified cascade scheme, as well as the baseline one from [13] does not require setting additional parameters. At the same time, the second modification of the method from [14] is based on an iterative procedure and demands the selection of the appropriate number of iterations to ensure the highest prediction accuracy.

Since the modeling was performed to predict two different indicators, the method was run twice with a maximum number of iterations of 100 to solve the heating and cooling loads prediction task in residential buildings. The results of these experiments are presented in Figures 3 and 4, respectively.

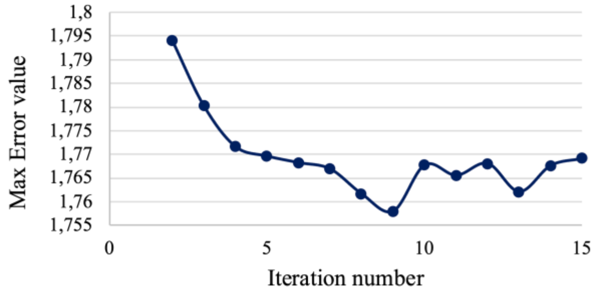

Figure 3. The optimal iteration number for the modification of the iterative cascade scheme (heating prediction)

As can be seen from Fig. 3, increasing the number of iterations of the modified cascade scheme provides an increase in accuracy for the heating load prediction task. However, this happens up to the 9th iteration, and then the accumulation of errors at each previous iteration provokes an increase in the overall error of the method (the maximum residual error). Therefore, the optimal number of iterations for solving the heating prediction task using modification 2 of the cascade scheme proposed in the paper will be 9.

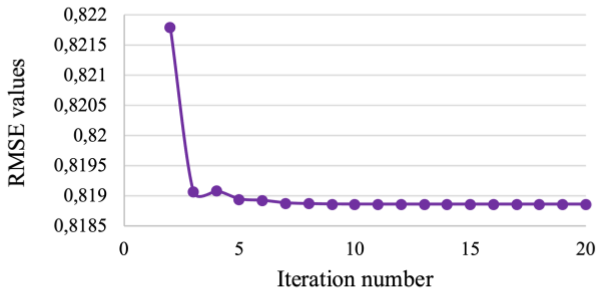

Figure 4. The optimal iteration number for the modification of the iterative cascade scheme (cooling prediction)

In the case of using a modified scheme with an iterative procedure for correcting predicted results to address the cooling load prediction task, the outcomes exhibited some variations. Similar to the prior scenario, an increase in the number of iterations of the method resulted in a reduction of the method error [16]. However, this trend persisted until a certain moment (10 iterations), beyond which the RMSE error reached the stage of saturation and remained constant with further iterations [17]. Hence, the optimal number of iterations for the modified cascade scheme in this case is determined to be 10.

C. Results

The evaluation of the effectiveness of both modified cascades took place using 9 performance indicators, including the number of iterations for the second modification, which includes an iterative procedure for specifying the prediction accuracy and the time of their training. The consolidated results for all indicators are presented in Table 1.

As can be seen from Table 1, both modifications, according to the coefficient of determination (R^2), demonstrate notably high prediction accuracy.

TABLE I. Results for both modified methods

Performance indicators	Methods			
	Modified basic cascade scheme (Modification 1)		*Modified iterative cascade scheme (Modification 2)*	
	Heating	*Cooling*	*Heating*	*Cooling*
MaxE	1.79	4.14	1.76	4.1
RMSE	0.41	0.83	0.38	0.82
MSE	0.17	0.69	0.15	0.67
MedAE	0.17	0.33	0.16	0.3
MAE	0.28	0.54	0.25	0.52
MAPE	0.2	0.02	0.1	0.02
R^2	0.998	0.992	0.999	0.993
Iterations number	-	-	9	10
Training time (seconds)	0.274	0.94	1.94	8.76

- means that this parameter does not apply to the method

However, the cascade scheme modification [14], incorporating an iterative procedure for parameter corrections in the regression model, resulted in a reduction of all errors considered in the paper compared to the first modification of the cascade scheme [13]. Only the time of its training increased significantly, which is explained both by the non-linear expansion of the input data and the number of iterations needed to achieve optimal results.

IV. Comparison and Discussion

The efficiency comparison of both modified cascade schemes was conducted using the baseline cascades from [13] and [14].

Table 2 compiles the results of both baseline cascade schemes and their respective modifications proposed in this paper when solving the tasks outlined in the work. The comparison is based on two efficiency indicators, MAE and MSE.

TABLE II. Comparison of both existing cascade schemes and their modifications

Performance indicators	Methods			
	Heating load prediction task		*Cooling load prediction task*	
	MAE	*MSE*	*MAE*	*MSE*
Basic Cascade from [13]	0.28	0.17	0.57	0.73
Iterative cascade from [14]	0.26	0.15	0.55	0.69
Modification of [13]	0.28	0.17	0.54	0.69
Modification of [14]	0.25	0.15	0.52	0.67

It should be noted that all other performance indicators used in the paper yield similar results, which will be discussed in detail.

When solving the heating load prediction task, Modification of [14] showed an almost 10% reduction in MAE error compared to Basic Cascade from [13] and more than 4% compared to Iterative cascade from [14].

Modeling of the operation of the modified methods for the cooling load prediction task showed that the Modification of [13], demonstrated a reduction of the MSE error by more than 5% compared to the Basic Cascade from [13], and the Modification of [14] by more than 3% compared to Iterative cascade from [14].

It should also be noted that the optimal results of the iteration procedure for the Modification of [14] were obtained with 10 iterations (cooling load prediction task). At the same time, optimal results for the Iterative cascade [14] were obtained when 24 iterations were used. However, taking into account the increase in dimensionality of the input data space for the second level of the cascade, the duration of its training has increased significantly. This is the main disadvantage of both modifications proposed in this paper.

If we compare the accuracy of the work of the best existing method for solving the stated task, the Gradient boosting regressor [15], then both the Basic and Modified cascade schemes demonstrated a significant increase in the predicted accuracy for the heating and cooling loads in residential buildings (Figs. 5, 6).

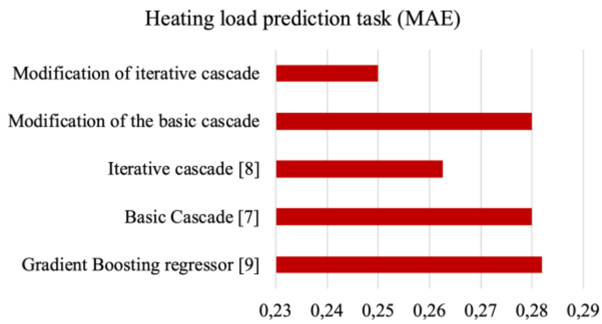

Figure 5. MAE values for all investigated methods (heating prediction task)

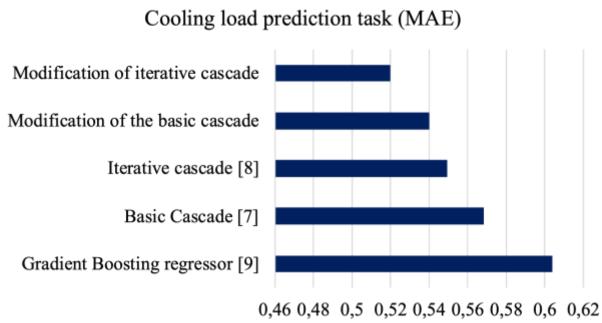

Figure 6. MAE values for all investigated methods (cooling prediction task)

However, as depicted in Fig. 5, the best results were obtained precisely for both cascade schemes modified in this paper. It is important to note that the proposed approach can be applied to improve the accuracy of diverse machine learning methods in various application areas [18], [19].

Conclusions

The present study addresses the issue of enhancing machine learning (ML)-based predictions for heating and cooling loads in residential buildings. The solution proposed in this paper builds upon an established ensemble method within machine learning, known for its superior accuracy in addressing the aforementioned problem. The authors have implemented improvements to enhance accuracy, focusing on cascade schemes.

Specifically, two two-step cascade correction schemes for the parameters of the regression model (employing a direct approach) were modified. This modification was achieved through the utilization of nonlinear expansion of inputs for the second machine learning method in the cascade, employing the Wiener polynomial of the second degree. Optimal operating parameters for these modifications were identified, demonstrating enhanced operational efficiency in comparison to both basic cascade schemes and the Gradient Boosting Regressor when applied to predict the heating and cooling loads in residential buildings.

The proposed approach outlined in this paper holds the potential to enhance the accuracy of diverse machine learning methods across various application domains. Furthermore, it can contribute to the establishment of a foundation for Resilient AI, particularly within the realm of civil engineering [20], [21].

Subsequent research endeavors will focus on refining an inverse approach to cascade approximation through rational fractions to achieve greater accuracy. In this context, the second machine learning method within the cascade will not only consider the initial independent variables but also incorporate the outcomes of the first machine learning method as an additional independent feature. This refined approach is anticipated to substantially elevate the prediction accuracy, applicable to both two-level and multi-level cascade schemes.

REFERENCES

[1] B. Gebreslassie, A. Kalam, and A. Zayegh, 'Design, Modeling of a Green Building Energy Optimized Efficient System', in *2023 7th International Conference on Green Energy and Applications (ICGEA)*, Singapore, Singapore: IEEE, Mar. 2023, pp. 185–191.

[2] D. Sciuto and A. A. Nacci, 'On How to Design Smart Energy-Efficient Buildings', in *2014 12th IEEE International Conference on Embedded and Ubiquitous Computing*, Milano, Italy: IEEE, Aug. 2014, pp. 205–208. doi: 10.1109/EUC.2014.37.

[3] S.-A. Mitoulis, D. V. Bompa, and S. Argyroudis, 'Sustainability and climate resilience metrics and trade-offs in transport infrastructure asset recovery', *Transportation Research Part D: Transport and Environment*, vol. 121, p. 103800, Aug. 2023, doi: 10.1016/j.trd.2023.103800.

[4] O. Pavliuk, T. Steclik, and P. Biernacki, 'The forecast of the AGV battery discharging via the machine learning methods', in *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan: IEEE, Dec. 2022, pp. 6315–6324. doi: 10.1109/BigData55660.2022.10020968.

[5] O. Pavliuk, R. Cupek, T. Steclik, M. Medykovskyy, and M. Drewniak, 'A Novel Methodology Based on a Deep Neural Network and Data Mining for Predicting the Segmental Voltage Drop in Automated Guided Vehicle Battery Cells', *Electronics*, vol. 12, no. 22, p. 4636, Nov. 2023, doi: 10.3390/electronics12224636.

[6] V. Kovtun, K. Grochla, V. Kharchenko, M. A. Haq, and A. Semenov, 'Stochastic forecasting of variable small data as a basis for analyzing an early stage of a cyber epidemic', *Sci Rep*, vol. 13, no. 1, p. 22810, Dec. 2023, doi: 10.1038/s41598-023-49007-2.

[7] I. Izonin, R. Tkachenko, S. A. Mitoulis, A. Faramarzi, I. Tsmots, and D. Mashtalir, 'Machine learning for predicting energy efficiency of buildings: a small data approach', *Procedia Computer Science*, vol. 231, pp. 72–77, 2024, doi: 10.1016/j.procs.2023.12.173.

[8] I. Krak, O. Barmak, E. Manziuk, and A. Kulias, 'Data Classification Based on the Features Reduction and Piecewise Linear Separation', in *Intelligent Computing and Optimization*, vol. 1072, P. Vasant, I. Zelinka, and G.-W. Weber, Eds., in Advances in Intelligent Systems and Computing, vol. 1072. , Cham: Springer International Publishing, 2020, pp. 282–289. doi: 10.1007/978-3-030-33585-4_28.

[9] M. Havryliuk, R. Kaminskyy, K. Yemets, and T. Lisovych, 'Interactive Information System for Automated Identification of Operator Personnel by Schulte Tables Based on Individual Time Series', in *Advances in Artificial Systems for Logistics Engineering III*, vol. 180, Z. Hu, Q. Zhang, and M. He, Eds., in Lecture Notes on Data Engineering and Communications Technologies, vol. 180. , Cham: Springer Nature Switzerland, 2023, pp. 372–381. doi: 10.1007/978-3-031-36115-9_34.

[10] V. Shymanskyi and V. Liaskovets, 'Cascade Model for Price and Time of Car Sales Prediction', *CEUR-WS.org*, vol. 3641, pp. 152–167, 2024.

[11] D. Chumachenko, P. Piletskiy, M. Sukhorukova, and T. Chumachenko, 'Predictive Model of Lyme Disease Epidemic Process Using Machine Learning Approach', *Applied Sciences*, vol. 12, no. 9, p. 4282, Apr. 2022, doi: 10.3390/app12094282.

[12] F. Geche, O. Mitsa, O. Mulesa, and P. Horvat, 'Synthesis of a Two Cascade Neural Network for Time Series Forecasting', in *2022 IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC)*, Kyiv, Ukraine: IEEE, Oct. 2022, pp. 1–5. doi: 10.1109/SAIC57818.2022.9922991.

[13] I. Izonin, R. Tkachenko, O. Shcherbii, O. Berezsky, I. Krak, and M. Oliinyk, 'An Approximation Cascade Scheme via Rational Fractions for Biomedical Data Analysis', in *2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT)*, Lviv, Ukraine: IEEE, Oct. 2023, pp. 1–4. doi: 10.1109/CSIT61576.2023.10324122.

[14] I. Izonin, 'An Ensemble Method for the Regression Model Parameter Adjustments: Direct Approach', *mcm-tech*, vol. 24, pp. 35–44, Dec. 2023, doi: 10.32626/2308-5916.2023-24.35-44.

[15] 'Energy Efficiency Model for Building'. Accessed: Jan. 29, 2024. [Online]. Available: https://kaggle.com/code/sasakitetsuya/energy-efficiency-model-for-building

[16] C. Ganguli, S. K. Shandilya, M. Nehrey, and M. Havryliuk, 'Adaptive Artificial Bee Colony Algorithm for Nature-Inspired Cyber Defense', *Systems*, vol. 11, no. 1, p. 27, Jan. 2023, doi: 10.3390/systems11010027.

[17] L. Mochurad and M. Yatskiv, 'Simulation of a Human Operator's Response to Stressors under Production Conditions', *CEUR-WS*, vol. 2753, pp. 156–169.

[18] L. Zomchak and V. Melnychuk, 'Creditworthiness of Individual Borrowers Forecasting with Machine Learning Methods', in *Advances in Artificial Systems for Medicine and Education VI*, vol. 159, Z. Hu, Z. Ye, and M. He, Eds., in Lecture Notes on Data Engineering and Communications Technologies, vol. 159. , Cham: Springer Nature Switzerland, 2023, pp. 553–561. doi: 10.1007/978-3-031-24468-1_50.

[19] L. Zomchak and I. Starchevska, 'Macroeconomic Determinants of Economic Development and Growth in Ukraine: Logistic Regression Analysis', in *Advances in Intelligent Systems, Computer Science and Digital Economics IV*, vol. 158, Z. Hu, Y. Wang, and M. He, Eds., in Lecture Notes on Data Engineering and Communications Technologies, vol. 158. , Cham: Springer Nature Switzerland, 2023, pp. 358–368. doi: 10.1007/978-3-031-24475-9_31.

[20] I. Linkov *et al.*, 'Toward Mission-Critical AI: Interpretable, Actionable, and Resilient AI', in *2023 15th International Conference on Cyber Conflict: Meeting Reality (CyCon)*, Tallinn, Estonia: IEEE, May 2023, pp. 181–197. doi: 10.23919/CyCon58705.2023.10181349.

[21] V. Kharchuk and I. Oleksiv, 'The Intellectual Structure of Sustainable Leadership Studies: Bibliometric Analysis', in *Advances in Intelligent Systems, Computer Science and Digital Economics IV*, vol. 158, Z. Hu, Y. Wang, and M. He, Eds., in Lecture Notes on Data Engineering and Communications Technologies, vol. 158. , Cham: Springer Nature Switzerland, 2023, pp. 430–442. doi: 10.1007/978-3-031-24475-9_37.