Supervised machine learning with regression for the IRT-T reactor cooling system

Maxim Kublinskiy*, *Nikita* Smolnikov, and *Artem* Naymushin[1]

Tomsk Polytechnic University, Nuclear Fuel Cycle Department, Lenina Ave 2, Tomsk, 634034, Russian Federation

Abstract. The purpose of this study is to create a machine learning model for the IRT-T reactor cooling system, which can estimate and predict the temperature difference in the secondary circuit. To do this, data was downloaded from the SCADA system, then an application was developed for converting and preprocessing this data. Then regression and classification models were constructed that evaluated the efficiency of the cooling system and its ability to predict changes in the temperature drop on heat exchangers. The main technical characteristics of the IRT-T reactor include a thermal power of 6 MW, the use of UO_2 nuclear fuel in an aluminum matrix with an enrichment of 90.1%, a coolant in the form of desalinated water, tubular square-section fuel rods with external cooling, the fuel element shell material is SAV-1 alloy, and the use of five 5 IRT-1000 type heat exchangers with a total area of heat exchange of 1000 m^2.

1 Introduction

The IRT-T reactor, located in Siberia and the Far East, is distinguished by its pool-type design, the interior of which is filled with demineralized water. This design performs three main functions: shielding, coolant and moderator. Various studies in the field of nuclear physics, radiobiology and neutron activation analysis are carried out in these regions, as well as various types of geochemical studies, production of radiopharmaceuticals and doped silicon are carried out.

The reactor cooling circuit consists of five surface-type heat exchangers, each of which has a heat transfer area of 200 m^2. The secondary circuit is cooled by ordinary water from an artesian well containing impurities, dirt and other elements that can contaminate the surface of heat exchangers. The heat exchangers are cleaned once a year, but this process is not optimized and is carried out according to the calendar schedule [1].

The relevance of this work lies in the importance of ensuring proper operating conditions for research nuclear facilities to ensure safety and the implementation of high-tech research.

The main objective of the project is to create a machine learning model for the RT cooling system, which will be used to analyze and predict temperature changes in the second circuit. So that was necessary to create special software for pre-processing of SCADA system data,

* Corresponding author: mkk4@tpu.u

 https://doi.org/10.1051/itmconf/20245903007

to analyze how preprocessing converters affect learning accuracy and to evaluate the optimal machine learning model based on Time Series Split and by Default Split (80/20) [2].

2 Materials and methods

2.1 IRT-T reactor cooling circuits

2.1.1 Primary circuit

The main components of the primary circuit - the main circulation circuit of the reactor are given as follows [1].

Reactor vessel with core and internal storage tank used for the full composition and isolation of nuclear materials.

Suction pipeline is the pipeline located before the pumps of the primary circuit.

External holding tank is the volume that reduces the activity of the components and allows burnable gases to decay.

Four main circulation pumps have the purpose to force the coolant to flow through the primary circuit.

Discharge pipeline is located after the pumps and heat exchangers and injects water into the vessel.

Emergency cooling pump has the same purpose as the main ones, but also works in the conditions without electricity.

Five heat exchangers have the intratubular space where the coolant of the primary circuit circulates and gives heat to secondary coolant.

Collectors to gather all the primary coolant in case of leakage.

Inlet and outlet pipelines of heat exchangers and pumps are used to introduce and intake the flow into the circuit.

Distribution box is used for correct distribution of the coolant among the volume of the vessel.

Natural circulation valves are used for emergency cooling when all the pumps are not in operation.

Drainage pipelines with shut-off valves used to drainage all the system if necessary.

2.1.2 Secondary circuit

The secondary circuit comprises the following primary elements [1].

Suction pipeline is the pipeline located before the pumps of the secondary circuit.

Four circulation pumps have the purpose to force the coolant to flow through the secondary circuit.

Discharge pipeline is located after the pumps and heat exchangers and injects water into the cooling tower.

Five heat exchangers have the in tubular space where the coolant of the secondary circuit circulates and takes heat from primary coolant.

Collectors to gather all the secondary coolant in case of leakage.

Inlet and outlet pipelines of heat exchangers and pumps are used to introduce and intake the flow into the circuit.

Biological protection cooling heat exchanger is used to decrease the temperature in the concrete around the reactor core by some tubes located around experimental channels.

Cooling tower has three sections with fans and nozzle to provide the sufficient heat removal from secondary coolant.

Main pipelines Ø 400 mm, Ø 600 mm are used for water flow pumping.

Valve switching chamber KPZ-1 allows personal to shut off some sections of cooing tower.

2.2 Machine learning model for the heat exchanger

Machine learning is a fundamental element within the realm of artificial intelligence, designed to construct analytical models through the process of learning from past data. The origins of artificial intelligence and ML can be dated back to the mid-20th century, when innovator Alan Turing introduced the concept of developing a "machine capable of learning from experience." Following years of continuous advancement and technological progress, ML has evolved into a potent tool with extensive applications across scientific inquiry and industry, notably excelling in extracting patterns from intricate, multi-dimensional datasets and exploring non-linear relationships [3].

Machine studies the relations between varieties and does in the different method classified in the method given in given in Figure 1.

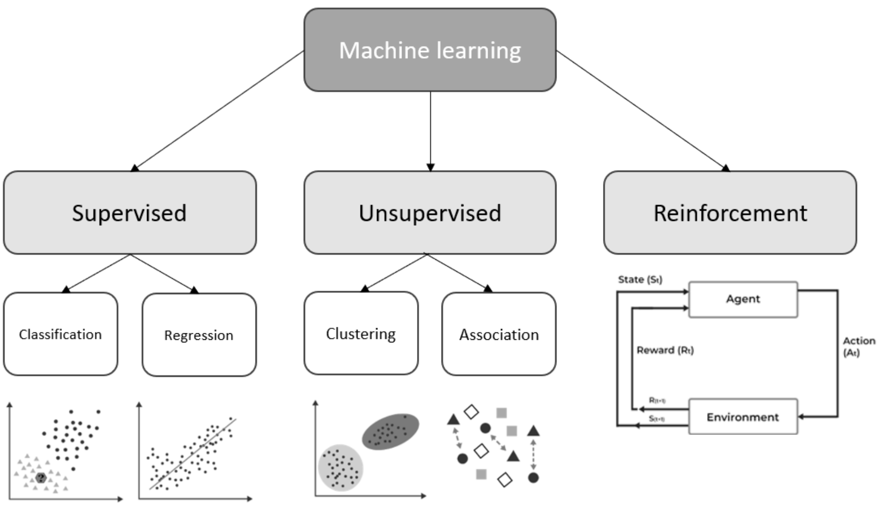

Fig. 1. The basic structure of machine learning.

Supervised learning represents a subdivision within both the fields of machine learning and artificial intelligence. Its distinguishing feature lies in its utilization of labeled datasets to instruct algorithms in accurately classifying data or predicting outcomes. The model adapts its parameters as input data is introduced, ultimately achieving a well-tailored fit during the cross-validation phase. Harnessing supervised learning can empower organizations to tackle a wide array of real-world challenges on a large scale, including tasks such as sorting spam into a separate folder within an inbox [4].

In data mining, supervised learning can be categorized into two types of problems: classification and regression.

Classification entails using an algorithm to accurately allocate test data into specific categories, thereby identifying distinct entities within the dataset and deriving conclusions regarding how these entities should be labeled or defined. Key classification algorithms include linear classifiers, support vector machines (SVM), decision trees, k-nearest neighbor, and random forest, each of which is elaborated upon in the following sections.

Regression, on the other hand, aims to comprehend the relationship between dependent and independent variables and is often employed to make projections, such as forecasting sales revenue for a given business. Widely used regression algorithms include linear regression, logistical regression, and polynomial regression.

2.2.1 Decision tree structure

Ehe initial dataset was obtained from the SimpleSCADA complex software. It is crucial to consider parameters such as temperature and flow rate, as they significantly impact the heat transfer coefficient [5]. The anticipated structure is outlined in Figure 2, which depicts the structure shaping the decision tree.

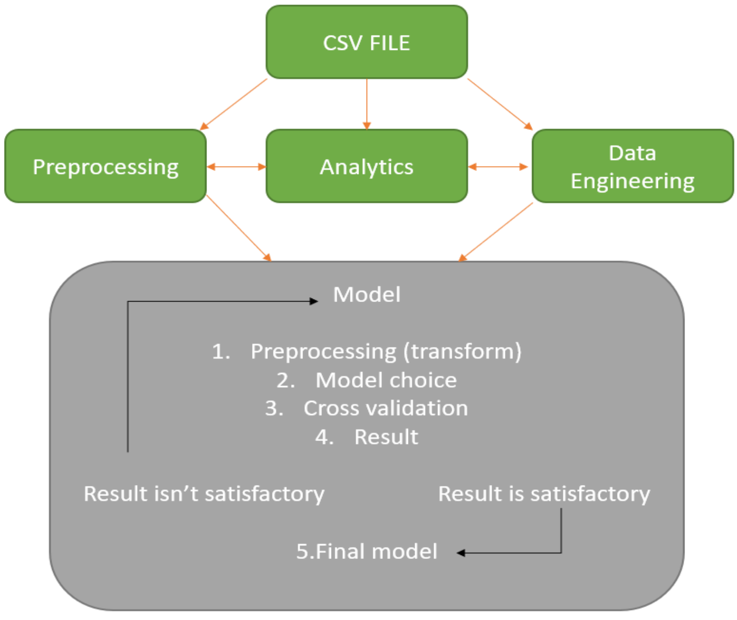

Fig. 2. Decision tree structure of the heat exchanger model.

The preprocessing, analytics, and data engineering are amalgamated into a single file and are elaborated upon more comprehensively in Subsection 2.3. The model encompasses the internal structure and internal data preprocessing tailored for modeling purposes. Various machine learning models need to be juxtaposed, and the most appropriate one must be selected. Subsequently, the final output needs to be assessed for its adequacy, and the ultimate model for predicting heat exchange must be chosen.

2.2.2 K-Fold cross-validation visualization

It's crucial to select the appropriate item for effective cross-validation to align with the model construction. Therefore, the data is divided into training and testing sets, a practice that helps prevent overfitting and allows for standardization across groups [6]. To achieve this, the Time Series Split utility is utilized, with a visual representation of the data analysis depicted in Figure 3.

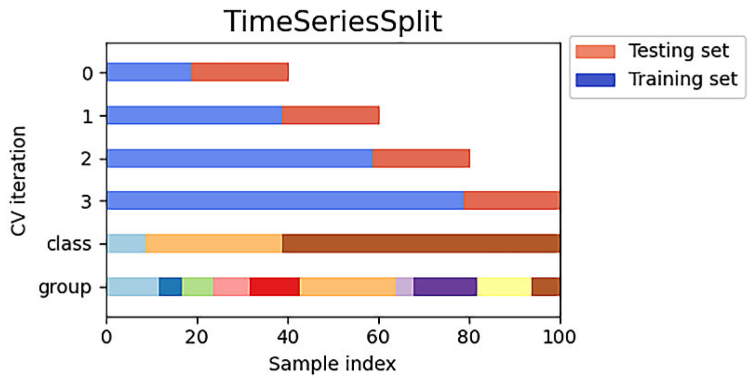

Fig. 3. Time Series Split visualization using K-Folds [5].

This approach, illustrated in Figure 4, represents the foundational process of cross-validation and underpins every teaching procedure in supervised learning.

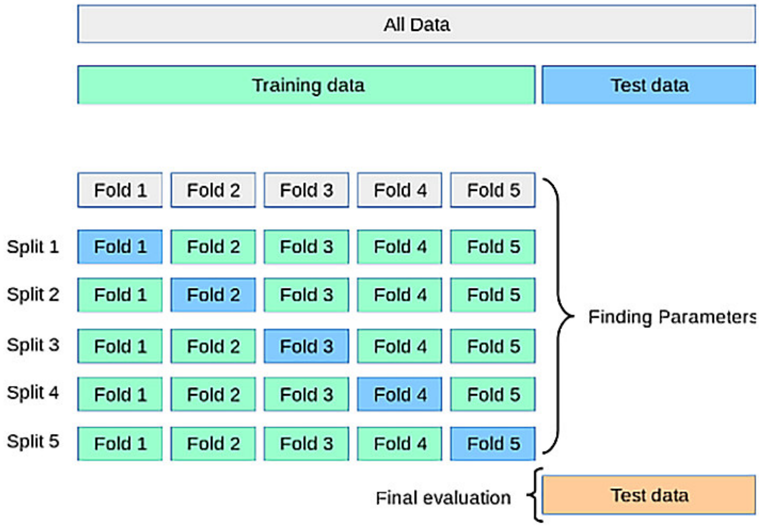

Fig 4. Test and trainings splitting for the dataset [6].

The process involves partitioning the voluminous dataset into distinct segments, with the machine systematically assimilating and testing each portion. The more comprehensive the training, the more optimal the results. The subsequent tests yield progressively more favourable outcomes, and the practical manifestation of this process is detailed in Section 3.

2.3 Building the general structure of the model

The initial step involved organizing a shared directory to facilitate collaborative work and exchange between the lead researcher and the supervisor. To achieve this, the master's laptop running Windows 10 Home version was enhanced by installing WSL and Ubuntu packages, enabling seamless operation on Linux within Microsoft Visual Studio Code without the need to switch between different operating systems.

Subsequently, it became apparent that machine learning is reliant on vast amounts of data, necessitating acquisition from a suitable database. In 2019, the IRT-T technological control system underwent an upgrade, incorporating the introduction of the SCADA system. This SCADA system not only furnishes the operator with essential information but also serves as an archive collector from the time of its implementation.

While working with the data, it became evident that the data units from were being recorded in the database in varying numerical formats and at different time intervals. In order to conduct a thorough analysis, it was essential for all data to conform to uniform numerical and time parameters. Consequently, a preprocessing.py file was created to address this requirement for future analysis.

3 Results

It is apparent that the water flow rate diminishes over time as a result of fouling and variations in pump operations. Although the heat exchange area is consistent for all heat exchangers, the flow rate for each differs, necessitating individual calculations for each unit. The water flow rate of first heat exchanger and first circulation pump current are shown on Figure 5 and Figure 6 respectively as an example of preprocessing and reconstruction of the data from the Subsection 2.3.

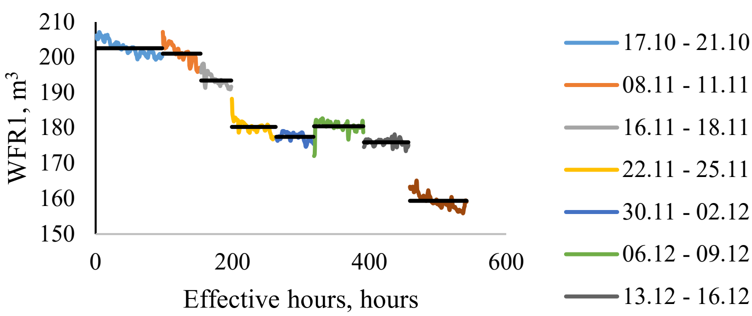

Fig. 5. The flow rate of the first heat exchanger decreasing through the time.

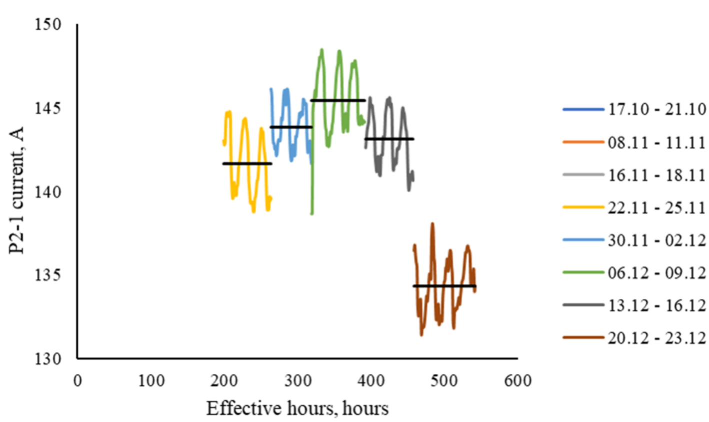

Fig. 6. The current of the first main circulation pump.

The data consists of direct values extracted from the .csv file and indirect values, such as thermal power and the heat transfer coefficient, derived from these source values. It is crucial to highlight that the thermal power value is divided equally by 4, corresponding to the number of operative heat exchangers, for ease of calculation (not reflective of actual conditions). The graphical portrayal depicting the variation of thermal power over time is depicted in Figure 7.

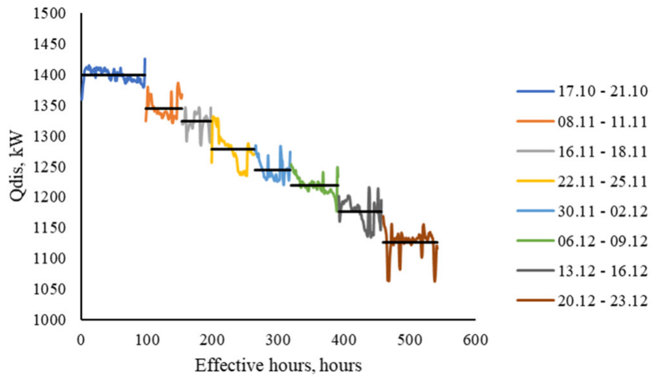

Fig.7. The average thermal power of each heat exchanger.

It is evidently apparent that the thermal power value decreases over time, a phenomenon largely attributed to the fouling of tubes within the heat exchanger. Additionally, it was imperative to ascertain the temperature difference and conduct the analysis and machine learning based on this variable, shown on Figure 8.

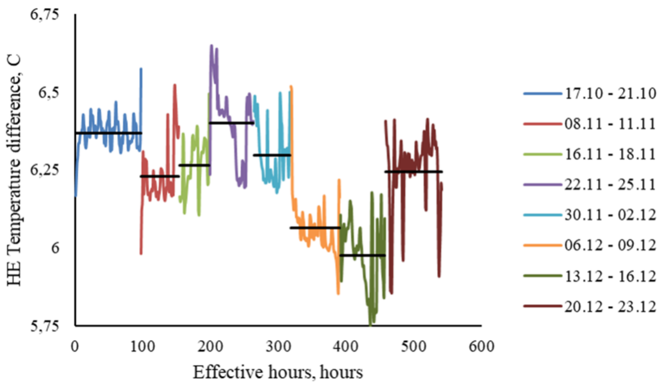

Fig. 8. The temperature difference for all heat exchangers in collectors.

Prior to delving into this model, several assumptions were established. It was determined that the most crucial parameters for analysis are inlet temperature before the heat exchangers, outside air temperature, cooling tower fan currents and temperature differences across the heat exchangers. To comprehensively study the model, it is imperative to construct a heatmap of these parameters, as depicted in Figure 9, to visualize the correlations among these variables.

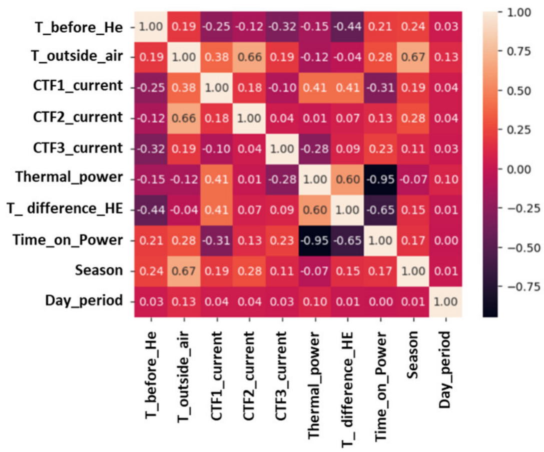

Fig. 9. The heat map for machine learning parameters.

As an illustration, the temperature difference increases during the operation of the first cooling tower fan (CTF1_ and decreases as the inlet temperature before the heat exchangers rises. The concept was to develop a Column Transformer function.

The preprocessed data required transformation into a format that is more conducive for machine learning. This involved taking the values from the dataset, calculating the average value (designated as 0), and identifying deviations from 0 in either a positive or negative direction. For instance, a small subset of the dataset pertaining to the temperature before the heat exchangers had an average temperature of approximately 22.8 °C, which was set as 0, with deviations established from this mean. This methodology serves to enhance the accuracy of calculations for all parameters. At present, the model is operating in a nearly binary fashion, allowing it to readily comprehend data deviations. The notion of partitioning the time series into segments is a straightforward concept. The supervisor is required to designate one segment as the training set (typically a larger subset) and the other as the test set. Subsequently, the process of cross-validation ensues, involving the selection of different partitions for training and testing, thereby enabling the model to undergo learning. A greater number of training iterations typically yields better results. Utilizing an 80/20 split for training/testing the data serves to smooth the dataset and appears to be optimal for machine learning. The subsequent step involved the selection of a suitable model, followed by a comparative analysis of errors, which is presented in Table 1.

Table 1. Comparative analysis and model choice.

Model	With transformers		Without transformers	
	Mean Error, °C	Root Error, °C	Mean Error, °C	Root Error, °C
KNeighbours	0.137+/- 0.019	0.168 +/- 0.024	0.187+/- 0.089	0.268 +/- 0.084
Multylayer perception	0.551 +/- 0.805	0.633 +/- 0.869	1.461 +/- 0.805	1.233 +/- 0.464
Gaussian processed regression	0.123 +/- 0.028	0.150 +/- 0.031	0.525 +/- 0.029	0.751 +/- 0.037
Gradient boost regression	0.118 +/- 0.019	0.146 +/- 0.023	0.124 +/- 0.016	0.152 +/- 0.019
Stochastic Gradien Descent with Warm Restarts	0.764 +/- 0.741	0.884 +/- 0.829	Algorithm crashed	Algorithm crashed

That is clearly visible that Gradient Boost Regressor shows the least error. 0.1 of °C is a perfect result for machine learning. Transformers is really important step because it decreases deviation. Default split (80/20) for training shows very good result. Using the simple square analysis for the whole Dataset interval such results were obtained. Prediction figure covers the real one on 0.90. So, the accuracy of prediction for the model is nearly 90 %. That is seen on Figure 10.

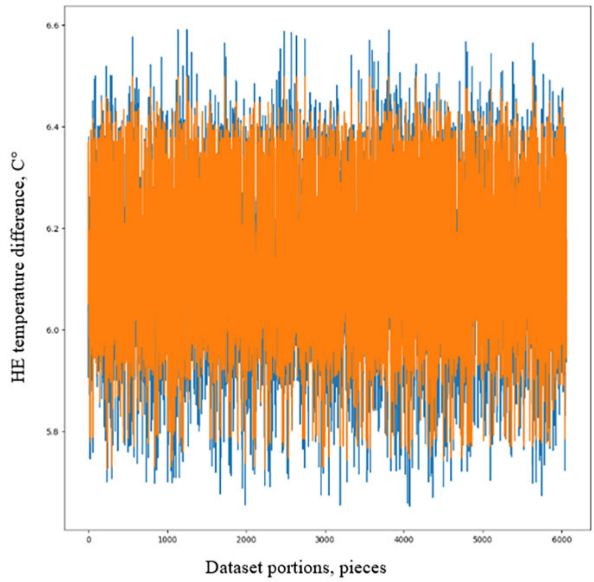

Fig. 10. Default Split (80/20).

The ultimate outcome is depicted through the graphs generated from the Time Series data shown in Figure 11, wherein the blue line represents the actual state and the black line illustrates the model's performance on unfamiliar data. It is worth noting that the model is capable of detecting the data peaks, with deviations from the actual state not exceeding 0.1 to 0.2 °C.

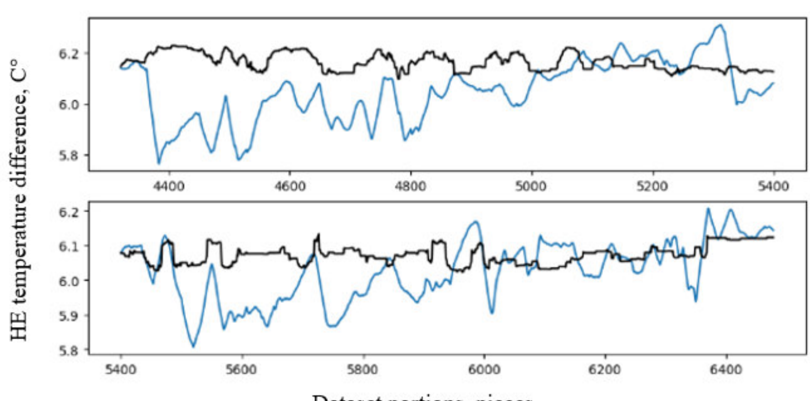

Fig. 11. Time Series Split.

This model version exhibits applicability in manufacturing. By incorporating additional data and training, it is possible to refine and reduce the observed deviation, representing a promising avenue for the advancement of this research.

4 Conclusions

In the course of the research a survey of various methods for machine learning modeling was conducted. For this particular case study, supervised learning with regression appeared to be the most effective approach. Also, the custom utility named CSV Refactorer was developed as a preprocessing tool for CSV files obtained from the Simple SCADA system. This software enables file reconstruction, column division, and differentiation of time intervals by one hour. To enhance the data's interpretability for the machine model, an assembly of transformers (One Hot Encoder + Standard Scaler) was utilized. This approach led to a 20% reduction in modeling errors and a consequent improvement in accuracy. An evaluation of different machine learning models was carried out. Based on the temperature difference data, the Gradient Boosting Regressor model exhibited the best results, achieving an accuracy of approximately. Utilizing this model for estimating personnel allocation over a month and week could yield a prediction accuracy of up to 90%.

References

1. M. Anikin, I. Lebedev, A. Naymushin, A. Ovsenev, N. Smolnikov, P. Khudoleev, *Laboratory practicum at the nuclear research reactor IRT-T. Part 1: educational and methodical manual* (2021)
2. S. Sokhina, S. Nemchenko, *Machine Learning, Machine learning methods*, in Proceedings of the V Internatinoal Scientific Conference : Modern science in the conditions of modernization processes : problems, realities, prospects, 30 April 2021, Ufa, Russia (2021)
3. Sh. Kawaguchi, Y. Sato, H. Nakayama, *Machine learning model for analyzing situations on programming learning,* in Proceedings of the III International Conferecne of Big Data Analysis, ICBDA, 12 March 2018, Shanghai, China (2018)
4. N. Wiranda, F. Sadikin, *Machine learning for security and security for Machine learning*, in Proceedings of IV International Conference on Information and Communications Technology, ICOIACT, 30 August 2021, Virtual, Online (2021)
5. B. Nayak, R. Ojha, P. Subbarao, V. Batth, *Machine learning finance: Application of Machine learning in collaborative filtering recommendations,* International Journal of Recent Technology and Engineering, 905-909 (2019)
6. P. Pico-Valencia, O. Vinueza-Celi, J. Holgardo-Terriza, *Bringing Machine learning predictive models based on Machine learning to non-technical users*, Advances in Intelligent Systems and Computing, 3-15 (2021)