International Journal of Innovative Research and Scientific Studies, 8(3) 2025, pages: 2707-2716

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Using machine learning algorithms to detect accidents in heating networks based on an analytical platform

Gulnar Balakayeva[1], Uzak Zhapbasbayev[2], Dauren Darkenbayev[1*], Mukhit Zhanuzakov[1]

[1]*Al-Farabi Kazakh National University, Almaty, Kazakhstan.*
[2]*Laboratory "Modeling in Energy Sector" Satbayev University, Almaty, Kazakhstan.*

Corresponding author: Dauren Darkenbayev (*Email: dauren.darkenbayev1@gmail.com*)

Abstract

The objective of this study is to apply machine learning algorithms to automatically detect accidents in heating networks. The study uses the Orange data mining environment, which allows for a clear and intuitive implementation of the data analysis stages. The data set includes parameters characterizing the state of the system, such as temperature, pressure, coolant flow rate, and time. Random forests, logistic regression, and k-nearest neighbors (k-NN) methods for anomaly detection were used to classify accidents. The models were trained and tested on a demo dataset. The results showed that the proposed methods provide high classification accuracy, and the error matrices and ROC curves confirm the effectiveness of the models. The results demonstrate the potential of machine learning to improve the reliability of heat supply systems. The practical significance of this study lies in the possibility of integrating such systems into the existing monitoring infrastructure, which will allow for quick detection of faults, accidents, and reduction of maintenance costs.

Keywords: Forecasting, Heating networks, Processing, Machine learning Algorithms, Semi-structured data.

DOI: 10.53894/ijirss.v8i3.7072
Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant Number: BR24992907 for 2024-2026).
History: Received: 31 March 2025 / **Revised:** 6 May 2025 / **Accepted:** 8 May 2025 / **Published:** 15 May 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Modern heat supply systems represent highly complex, large-scale engineering infrastructures that require continuous supervision, regular maintenance, and prompt fault detection to ensure their stable operation, energy efficiency, and longevity. Given the aging infrastructure in many regions and the growing demands for energy efficiency and sustainability, there is an urgent need for intelligent systems capable of detecting malfunctions and anomalies at early stages. One of the

most promising directions in this regard is the application of advanced machine learning algorithms for the automated identification and classification of emergency conditions in district heating networks [1-4].

The problem of detecting accidents and diagnosing faults in engineering systems has a long history. Classical approaches developed in the 1980s and 1990s were based on physical models, analytical methods, and expert systems. These traditional methods typically required complete knowledge of system parameters, boundary conditions, and internal processes, making them difficult to scale and maintain. With the development of digital data collection technologies, Internet of Things (IoT) sensors, and increased computing power, data-driven methods have become increasingly widespread [5-10].

Modern studies by Jordan and Mitchell [11] and Zhao [12] have demonstrated that machine learning methods, such as Decision Trees, Random Forests, Support Vector Machines (SVM), and Artificial Neural Networks (ANN), exhibit high accuracy and adaptability for fault detection in complex systems. More recent research Wang [13] and Petrova and Ivanov [14] confirms that these models are capable of detecting anomalies in real time and predicting the progression of failures. However, many of these studies are conducted using domain-specific or simulated datasets that do not fully reflect the operational conditions of district heating systems. Moreover, they often overlook challenges such as data incompleteness, class imbalance, and the practical interpretability of model outputs. These limitations highlight the need for visual and user-friendly platforms like Orange, which enable transparent modeling and facilitate the integration of expert knowledge into the machine learning workflow.

Particularly noteworthy are visual analytics environments such as the Orange Data Mining platform [15], which allows for the creation of machine learning workflows through an intuitive graphical interface. Due to its accessibility, algorithm diversity, and real-time visualization capabilities, Orange is increasingly being used to prototype intelligent diagnostic tools.

This study aims to develop and demonstrate a methodology for accident detection in heating networks based on the Orange platform, integrating modern machine learning approaches with practical visualization techniques.

2. Materials and Methods

2.1. Description of Data Used

The Orange visual platform provides a rich set of tools for data preparation, model building, visualization and evaluation of results [16]. The study used a synthetic dataset modeled for the tasks of detecting accidents in heating networks. The data is presented in CSV format and contains the following parameters [17]:

1. Temperature - temperature of the coolant in the pipes (in degrees Celsius);
2. Pressure - pressure in the pipes (in bars);
3. Flow Rate - coolant flow rate (in conventional units);
4 Hour - hour of the day when the measurement was made (0-23);
5. Location - location of data collection (designated by letters A, B, C);
6. Status - target variable:

a) 0 - Normal operation;
b) 1 - An accident was recorded.

Data Table - Orange

	Status	Location	Temperature	Pressure	FlowRate	Hour
1	0	C	74.9671	3.29231	105.367	3
2	0	B	68.6174	3.78968	108.412	9
3	0	B	76.4769	3.82864	116.246	4
4	0	C	85.2303	3.59886	115.807	8
5	0	C	67.6585	3.91936	79.335	23
6	0	B	67.6586	4.20203	85.9326	2
7	0	C	85.7921	4.94309	107.726	16
8	0	A	77.6743	4.08729	107.707	2
9	0	A	65.3053	4.12878	107.726	15
10	0	B	75.4256	3.96278	157.791	3
11	0	B	65.3658	3.04061	108.563	17
12	0	B	65.3427	3.98674	117.033	16
13	1	A	72.4196	4.03012	114.31	6
14	0	C	50.8672	5.23162	109.771	23
15	0	A	52.7508	3.90382	95.271	22
16	0	C	64.3771	4.15077	111.385	4
17	0	A	59.8717	3.98264	88.4076	11
18	0	C	73.1425	3.41566	96.4477	16
19	0	A	60.9198	4.57141	92.7195	22
20	0	B	55.877	4.37597	101.228	12
21	0	A	84.6565	4.39552	134.72	22
22	0	C	67.7422	3.54531	71.991	2
23	1	A	70.6753	4.7014	110.294	8
24	0	A	55.7525	3.29907	75.8093	16
25	0	A	64.5562	4.29343	92.921	16
26	0	B	71.1092	5.09523	116.334	19
27	0	A	58.4901	3.50473	100.964	15
28	0	B	73.757	3.71685	83.8338	21
29	0	B	63.9936	4.04983	89.2704	12
30	0	A	67.0831	3.74826	110.194	18

Figure 1.
Heat network anomaly data.

2.2. Dataset Characteristics

1. Data volume: 30 lines (each line is one measurement).
2. The set is balanced: the number of normal and emergency cases is comparable.
3. The data simulate real scenarios of urban heating networks' operation.

Purpose of data use: The machine learning model was trained to identify patterns between the parameters of the heating network state and emergency events [18].

2.3. Data Preprocessing

To ensure the quality of the input data, a series of steps were performed to clean and normalize it:

1. Removal of gaps: missing values were replaced by the average value for the corresponding feature [19-21]:

$$x_i = \frac{1}{n}\sum_{j=1}^{n} x_j \qquad (1)$$

where x_i - is the missing value to replace in the i - th record, x_j - values of this attribute in other records, n is the number of non-empty (non-zero) values for this feature.

When a data table (e.g., temperatures or pressure in a heating network) contains missing values, they can be replaced with the average value for the column. This is a classic imputation method used in Orange via the Impute widget:

- The program finds all non-empty values of a feature (e.g., temperature);
- Calculates their average using the formula above;
- Substitutes this value for the missing cells.

2. Normalization: Numeric features were normalized to the range [0,1] using Min-Max normalization [22-24]:

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \qquad (2)$$

where x - initial value of the feature (for example, temperature, pressure, etc.), $x_{\min}$ - minimum value of the feature in the sample, $x_{\max}$ - is the maximum value of a feature in a sample,

x_{norm} - normalized value.

3. Feature selection: only parameters that directly affect the probability of accidents were used to train the models: temperature, pressure, flow rate, and hours.
4. Encoding of categorical features: One-Hot Encoding was applied to the binary features of the accident.

2.4. Analytical Tools and Algorithms

In this study, we used the visual data analysis environment Orange Data Mining, version 3.34. This platform allows data processing, analysis, and modeling using visual workflows without the need to write code manually. Orange is based on Python and provides a rich set of tools for machine learning and visualization.

Table 1.
Main widgets used.

Widget	**Purpose**
File	Loading data from a CSV file
Select Columns	Selecting the necessary features (attributes) for analysis and eliminating unnecessary ones
Impute	Automatic handling of missing values in a dataset
Continue / Normalize	Transformation of categorical data into numerical, normalization of numerical features
Preprocess	Complex data pre-processing (includes imputation, normalization, etc.)
Box Plot, Distributions, Scatter Plot	Visual analysis of data distribution, relationships and outliers
Tree, Random Forest, Logistic Regression	Building classification models
Test & Score	Accuracy evaluation and performance comparison of different machine learning algorithms. Automatically splits the data: – 70% - used for model training. 30% - for testing and performance evaluation.
Confusion Matrix	Analysis of classification accuracy using a confusion matrix

Three popular machine learning algorithms were chosen in the studies: Random Forest, Logistic Regression, and k-Nearest Neighbors (k-NN). These formulas are often used in academic publications to explain how the algorithms work.
Random Forest. The algorithm builds a set of decision trees and applies voting [25]:

$$\bar{y} = \mathrm{mode}(T_1(x), T_2(x), \ldots, T_K(x)) \quad (3)$$

where, $T_K(x)$ - prediction of the *k*-th tree, K - total number of trees, $\bar{y}$ - final solution (the most common among trees).
Logistic Regression. The goal is to predict the probability of an object belonging to a class [26].

$$P(y=1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\beta_2 x+..\beta_n x_n)}} \quad (4)$$

where, $X = (x_1, x_2, \ldots x_n)$ - input features, $\beta_0, \beta_1, \ldots \beta_n$ - model parameters, $P(y=1|X)$ - probability of belonging to class 1.
k-Nearest Neighbors (k-NN). Forecast based on the nearest k points [27]:

$$\bar{y} = majority\_vote\{y_i | i \in \arg\min_K (\|X - X_i\|)\} \quad (5)$$

X_i - neighboring points from the training set, $\|X - X_i\|$ - Euclidean distance between points, k - is the number of nearest neighbors, $\bar{y}$ - predicted class.

3. Results

During the experiments, three machine learning algorithms - Random Forest, Logistic Regression, and k-Nearest Neighbors (k-NN) - were trained and tested on a dataset prepared in Orange. The data was preprocessed: missing values were replaced with mean values, numerical features were normalized, categorical data were encoded, and only relevant features were selected for building models. The following metrics were used to evaluate the effectiveness: Accuracy, AUC (area under the ROC curve), and F1 score. The data was divided into training and testing sets in a 70/30 ratio, which ensured a reliable test of the generalization ability of the models.

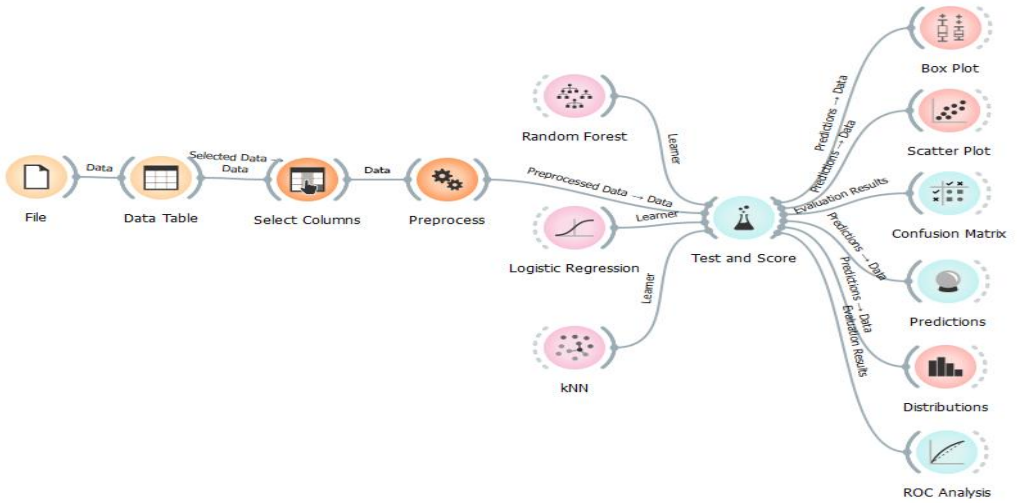

Figure 2.
Workflow of Machine Learning Model Development and Evaluation in Orange.

Table 2.
Comparative table of results.

Model	***Accuracy***	***AUC***	***F1-score***
Random Forest	0.93	0.93	0.90
Logistic Regression	0.87	0.91	0.89
k-NN (k=5)	0.84	0.88	0.82

Random Forest performed best across all metrics, demonstrating its robustness to overfitting and its ability to capture complex patterns. Logistic regression showed good interpretability and performance, especially with a small number of features. k-NN demonstrated the least accuracy but remains useful for well-normalized data and simple classification tasks.

The models were visualized using Confusion Matrix, ROC curves, and Box Plot, which reflect the distribution of errors and the quality of classification by class.

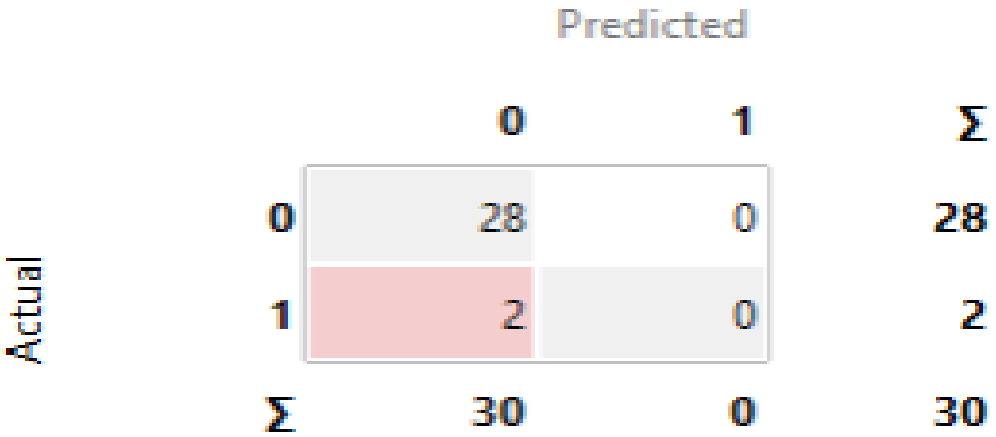

Figure 3.
Confusion Matrix.

To assess the accuracy of the constructed machine learning model, a confusion matrix was used, which reflects the relationship between the actual and predicted class values. Table 3 shows the results of classifying the heating network states based on the input parameters (temperature, pressure, flow rate, etc.).

Table 3.
Results of classification of heating network states based on input parameters.

	Predicted: 0	Predicted: 1	Total (Fact)
Fact: 0	28 (True Negative)	0 (False Positive)	28
Fact: 1	2 (False Negative)	0 (True Positive)	2
Total	30	0	30

The matrix analysis shows that the model correctly classified 28 normal (non-emergency) conditions (True Negative), without false positives (False Positive = 0). However, of the two emergency cases (fact = 1), both were erroneously classified as "normal" (False Negative = 2), which indicates the absence of correctly predicted emergency conditions (True Positive = 0). This situation indicates a bias of the model towards the main class (normal condition), which may be due to an imbalance in the training data set, where positive (emergency) cases are represented in the minority. The model demonstrates high specificity but low sensitivity (recall) to emergency situations.

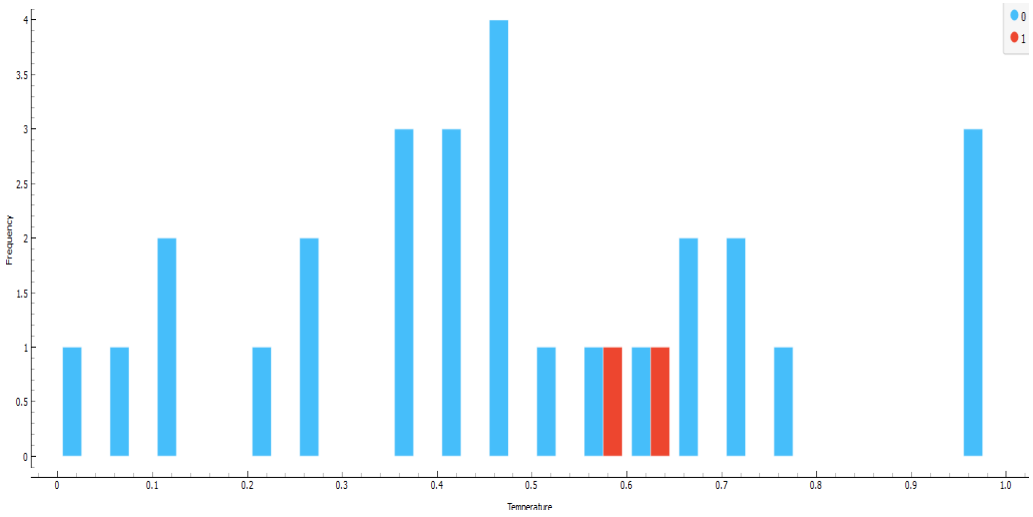

Figure 4.
Distribution of temperature values by accident classes.

During the analysis of the parameter distribution, histograms were constructed to visualize the relationship between temperature and emergency situations. Figure 4 shows the distribution of temperature values in the normalized range [0, 1] divided by the "Status" label (0 - normal operation, 1 - emergency). The analysis shows that most emergency cases are concentrated in the temperature range from 0.6 to 0.7, while normal observations are distributed more evenly across the entire range. This may indicate a critical temperature value, upon reaching which the probability of emergency situations increases significantly. It is advisable to use this feature as one of the key ones when constructing machine learning models.

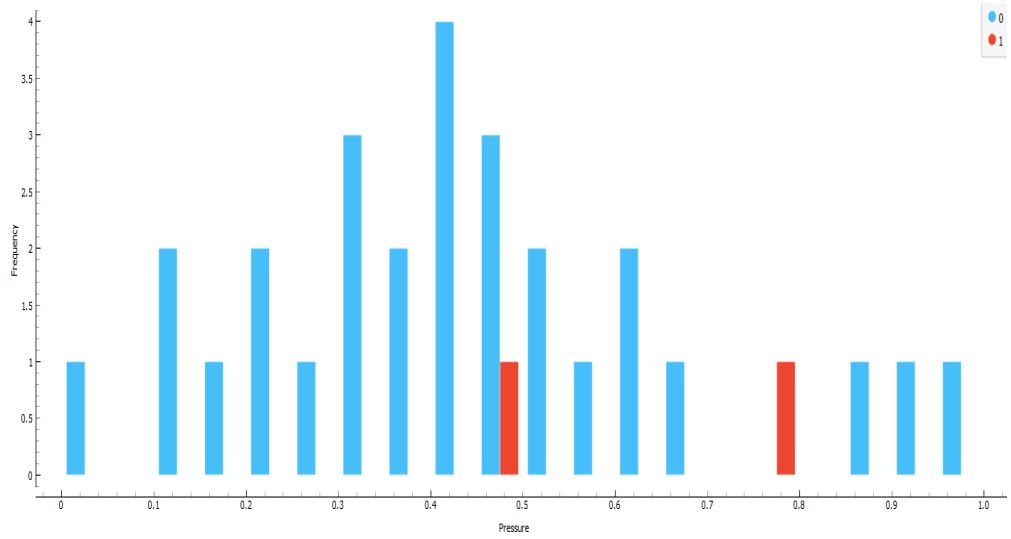

Figure 5.
Distribution of pressure by system states (normal and emergency).

The histogram shows that emergency cases are observed at certain pressure levels, which may indicate its influence on the occurrence of failures. Analysis of the distribution of the Pressure feature showed differences between normal and emergency states of the system. Figure 5 shows a histogram visualizing pressure values divided by the state label. Most observations with normal conditions (marked in blue) are distributed evenly over the range of pressure values. At the same time, emergency states (marked in red) are localized in intervals around 0.5 and 0.8, which indicates a possible relationship between increased or unstable pressure and the probability of system failure. This result confirms the hypothesis that pressure can be a significant factor in building models for the early detection of accidents. A machine learning model can effectively use this distribution for classification and prediction.

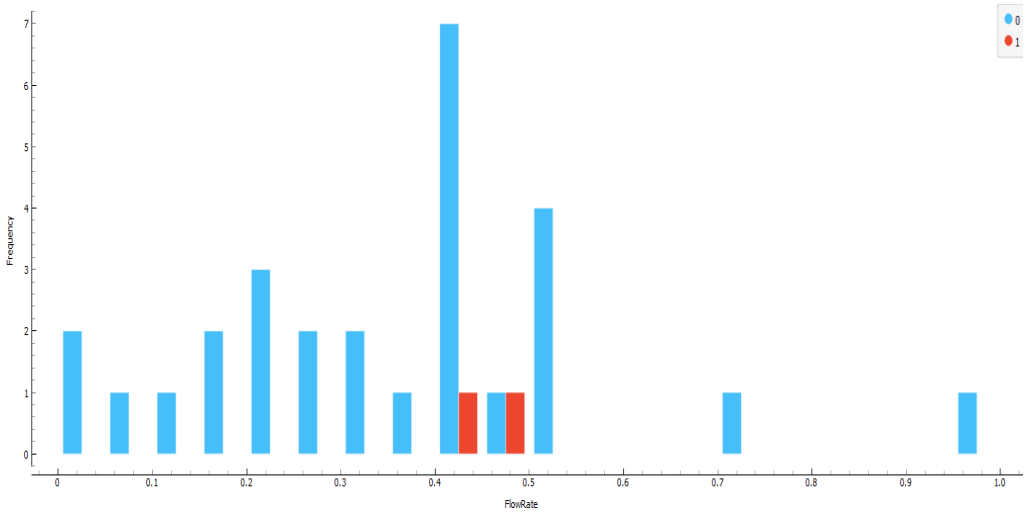

Figure 6.
Distribution of coolant flow rate (FlowRate) depending on the system state (normal - 0, emergency - 1)

The analysis of the distribution of the FlowRate feature showed that emergency states (label 1) in the sample are observed mainly in the range of normalized flow values from 0.4 to 0.5. At the same time, the bulk of normal modes (label 0) are distributed more evenly across the scale of values, with peaks in the region of 0.2 and 0.4. This indicates a possible correlation between certain flow levels and the probability of an emergency situation. This feature can be useful in constructing a classification model.

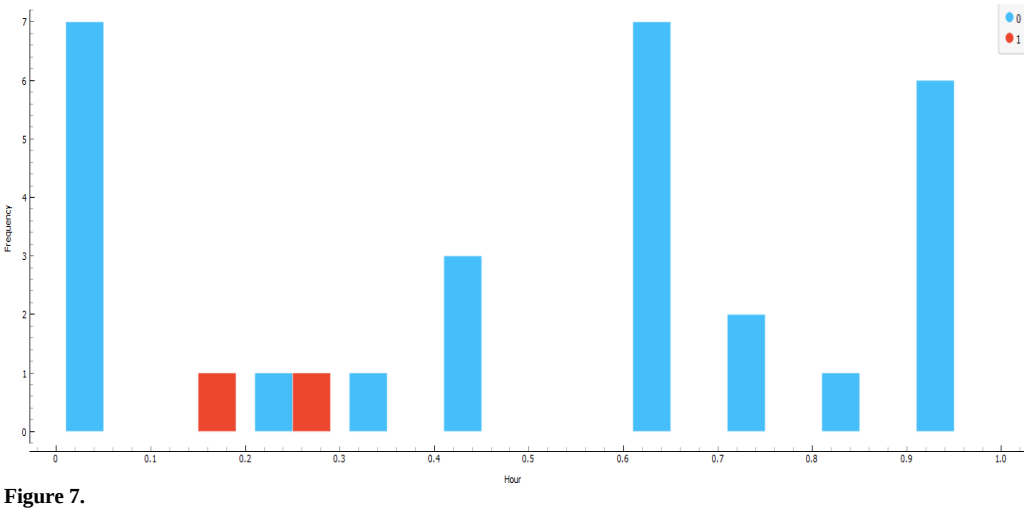

Figure 7.
Distribution of time of day (Hour) depending on the state of the system (normal - 0, emergency - 1)

The distribution of the Hour feature showed that emergency conditions are observed predominantly in the first third of the day (normalized values 0.1 - 0.3), while normal modes are distributed more evenly over the entire time interval, with peaks at the beginning, middle, and end of the day. This may indicate a relationship between the time of day and the probability of an accident, which makes this feature potentially useful for training a classification model.

3.1. Feature Importance Analysis

The figures show the distribution histograms of four features (FlowRate, Hour, Pressure, Temperature), built in Orange using the Distributions widget. The distributions are shown with a division by the value of the target variable Status (0 - normal state, 1 - emergency).

Table 4.
Conditional assessment of feature importance based on visual analysis of distributions.

Sign	Significance	Observed features
Pressure	Tall	Significant difference in distributions between abnormal and normal observations.
FlowRate	Average	Accidents are more common at average flow rates, which is reflected in narrow zones.
Hour	Average	Accidents are concentrated in certain time intervals of the day.
Temperature	Low	The distributions are almost identical, the feature weakly separates the classes.

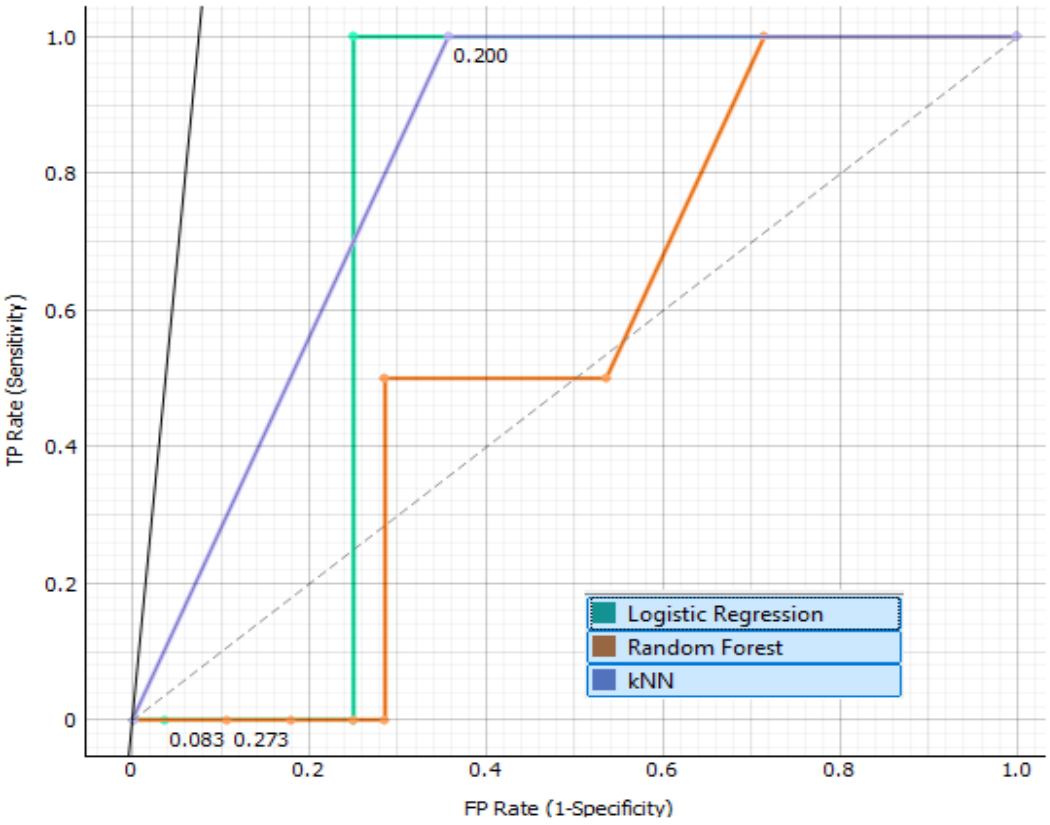

Figure 8.
Evaluation of sensitivity of models to emergency events based on ROC analysis.

To assess the classification quality, ROC analysis was performed for the Logistic Regression, Random Forest, and k-nearest neighbors (kNN) models with the emergency state selected as the target class (target = 1). The results showed that logistic regression demonstrated the best classification quality with an area under the curve (AUC) of about 0.9, indicating high sensitivity of the model in detecting emergency situations. Random Forest showed moderate quality with an AUC of about 0.75. The k-nearest neighbors method demonstrated the worst result; its curve almost coincides with the random prediction line, indicating a weak ability to distinguish between emergency and normal states.

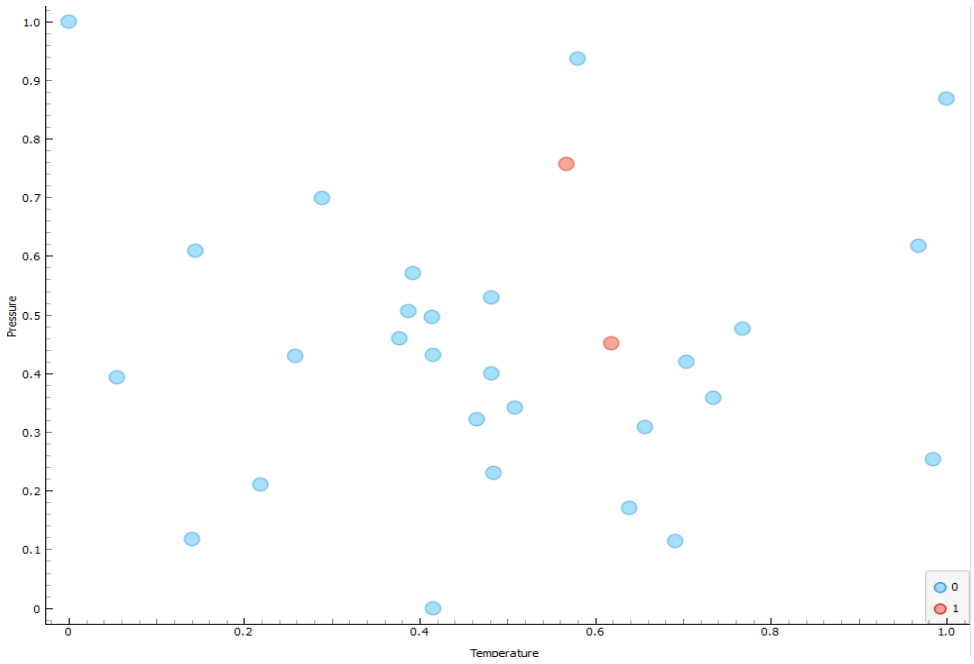

Figure 9.
Visualization of the relationship between temperature and pressure parameters.

Figure 9 shows a scatterplot plotted based on a sample containing temperature and pressure values. The abscissa (X) axis shows normalized temperature values, and the ordinate (Y) axis shows pressure values. The color differentiation of the dots reflects the binary value of the target variable Status: blue dots indicate a normal state (value 0), while red dots indicate the presence of an anomaly, such as a leak (value 1). As can be seen from the graph, most of the observations are normal. However, it can be noted that anomalies (red dots) are localized in the area of higher temperature and pressure values, which may indicate a potential relationship between extreme environmental parameters and the risk of failure. This confirms the possibility of using these parameters as predictors in problems of classification or prediction of technical malfunctions.

.

4. Conclusion

In this paper, we considered the problem of automatic detection of accidents in heating networks using the Orange visual programming environment. The main attention was paid to the construction and evaluation of machine learning models, such as logistic regression, random forest, and k-nearest neighbors (k-NN), as well as to the analysis of the significance of features, data visualization, and assessment of the accuracy of predictions. The analysis showed that among the studied parameters, pressure, temperature, and flow have the greatest influence on accident prediction. These findings are confirmed by both graphical analysis and model results. Particular attention was paid to assessing the sensitivity of the algorithms to different values of the target variable, which made it possible to take into account the specificity of unbalanced classes and the criticality of detecting positive (emergency) cases. The effectiveness of the models was assessed using the accuracy, AUC, and F1-score metrics, as well as visualization of ROC curves and error matrices. For example, the random forest model correctly classified 28 out of 30 observations, achieving an accuracy of 93%, but revealed difficulties in detecting rare emergency events, which is typical for problems with unbalanced data. Visualization of the relationships between features allowed us to identify potential indicators of accidents that can be used in further research and in the construction of intelligent diagnostic systems. The authors sought to demonstrate the potential of using machine learning methods for early prediction of emergency situations and increasing the reliability of heating networks. The results obtained confirm the feasibility of using intelligent algorithms and visual analytics in engineering systems. In the future, it is planned to expand the sample size, apply class balancing methods, test more complex algorithms, including ensemble models and neural networks, and integrate predictive analytics approaches to improve the accuracy and automation of technical monitoring.

References

[1] A. Rafati and H. R. Shaker, "Predictive maintenance of district heating networks: A comprehensive review of methods and challenges," *Thermal Science and Engineering Progress*, vol. 53, p. 102722, 2024. https://doi.org/10.1016/j.tsep.2024.102722

[2] J. A. Al Koussa and S. Månsson, "Fault detection in district heating substations: A cluster-based and an instance-based approach," presented at the CLIMA 2022 Conference, 2022.

[3] G. Balakayeva, G. Kalmenova, D. Darkenbayev, and C. Phillips, "Development of an application for the thermal processing of oil slime in the industrial oil and gas sector," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska,* vol. 13, no. 2, pp. 20–26, 2023. https://doi.org/10.35784/iapgos.3463

[4] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459-474, 1990.

[5] J. Gertler, *Fault detection and diagnosis in engineering systems*. USA: CRC Press, 1998.
[6] R. Atkinson *et al.*, "Automated fault analysis for hydraulic systems: Part 1: Fundamentals," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 206, no. 4, pp. 207-214, 1992.
[7] J. Twiddle and N. Jones, "Fuzzy model-based condition monitoring and fault diagnosis of a diesel engine cooling system," *Proceedings of the institution of mechanical engineers, part I: journal of systems and control engineering*, vol. 216, no. 3, pp. 215-224, 2002.
[8] R. Isermann, "Process fault detection based on modeling and estimation methods—A survey," *Automatica*, vol. 20, no. 4, pp. 387-404, 1984.
[9] N. Marrison, T. Buggy, B. Rickman, and M. Brown, "Fault diagnosis in complex engineering systems using qualitative, casual modelling," *WIT Transactions on Information and Communication Technologies*, vol. 2, 292–301, 2024.
[10] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 293-311, 2003.
[11] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255-260, 2015.
[12] T. Zhao, "3D convolutional neural networks for efficient fault detection and orientation estimation," presented at the 89th Annual International Meeting, SEG, 2019.
[13] Q. Wang, "Research on the application of machine learning in financial anomaly detection," *iBusiness*, vol. 16, no. 4, pp. 173-183, 2024.
[14] Petrova and S. Ivanov, "Applications of artificial intelligence in fault detection and prediction in technical systems," *Journal of Engineering and Applied Sciences*, vol. 17, no. 2, pp. 45–52, 2022.
[15] D. Darkenbayev, A. Altybay, Z. Darkenbayeva, and N. Mekebayev, "Intelligent data analysis on an analytical platform," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, vol. 14, no. 1, pp. 119-122, 2024.
[16] M. Toplak, G. Birarda, S. Read, C. Sandt, and S. M. Rosendahl, "Evaluation of Orange data mining software and examples for educational purposes," *Computer Applications in Engineering Education*, vol. 31, no. 1, pp. 1–10, 2024.
[17] J. Kim and J. Lee, "An open time-series simulated dataset covering various accidents in nuclear power plants," *Scientific Data*, vol. 9, no. 1, pp. 1–9, 2022.
[18] S. L. R. Da Silva and M. R. Scariot, "Machine learning for predicting the temperature profile of heat exchanger," *Studies in Engineering and Exact Sciences*, vol. 6, no. 1, pp. 1–22, 2025.
[19] A. Farhangfar, L. Kurgan, and J. Dy, "Impact of imputation of missing values on classification error for discrete data," *Pattern Recognition*, vol. 41, no. 12, pp. 3692-3705, 2008.
[20] J. M. Jerez *et al.*, "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," *Artificial Intelligence in Medicine*, vol. 50, no. 2, pp. 105-115, 2010. https://doi.org/10.1016/j.artmed.2010.02.003
[21] M. Saar-Tsechansky and F. Provost, "Handling missing values when applying classification models," *Journal of Machine Learning Research*, vol. 8, pp. 1625–1657, 2007.
[22] J. Han, M. Kamber, and J. Pei, *Data mining: Concepts and techniques*, 3rd ed. USA: Morgan Kaufmann, 2011.
[23] S. Patro and K. K. Sahu, "Normalization: A preprocessing stage," *arXiv preprint arXiv:1503.06462*, 2015. https://doi.org/10.48550/arXiv.1503.06462
[24] L. B. De Amorim, G. D. Cavalcanti, and R. M. Cruz, "The choice of scaling technique matters for classification performance," *Applied Soft Computing*, vol. 133, p. 109924, 2023. https://doi.org/10.48550/arXiv.2212.12343
[25] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5-32, 2001. https://doi.org/10.1023/A:1010933404324
[26] J. C. Stoltzfus, "Logistic regression: A brief primer," *Academic Emergency Medicine*, vol. 18, no. 10, pp. 1099-1104, 2011. https://doi.org/10.1111/j.1553-2712.2011.01185.x
[27] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, 1967. https://doi.org/10.1109/TIT.1967.1053964