International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 6 (2020), pp. 1243-1248

Creation of Forecast Algorithm for Networking Hardware Malfunction in the context of small number of breakdowns

A.A.Myrzatay[1] and L.G.Rzayeva[2]

[1]Second-year doctoral student of the Department of "System Analysis and Management" of the Faculty of "Information Technology" Eurasian National University. L.N. Gumilyov, Nur-Sultan, Republic of Kazakhstan.

[2]Dr. PhD, acting Associate Professor, Department of System Analysis and Management, Faculty of Information Technology, Eurasian National University. L.N. Gumilyov, Nur-Sultan, Republic of Kazakhstan.

ORCID: 0000-0002-5339-2437 (A.A.Myrzatay)

Abstract

Constant improvement of information technology leads to expansion of sizes of local networks of enterprises and issues connected to its organization and monitoring. Considering that every minute a huge amount of heterogeneous data is generated within the network, operating a computer network LAN becomes a nontrivial task. Since a computer network consists of a wide variety of devices (switches, servers, routers, computers etc.), and each of these devices utilizes a high number of application software, which is a source of immense amount of information, the forecast based on processing and analysis of this data would be best made by an algorithm of machine learning. The presence of a sufficiently accurate prediction of malfunction will be able to ensure high reliability, safety and economic efficiency of operating the equipment of local area networks segments.

This article proposes a model for predicting equipment failures in a particular case based on the Random Forest machine learning algorithm. The key stages of building and tuning the model are considered. The model includes several sub-models predicting equipment failure using actual and predicted readings from sensors.

Keywords: predicting failures, machine learning, Random Forest, decision tree, model.

INTRODUCTION

Constant improvement of information technology leads to expansion of sizes of local networks of enterprises and issues connected to its organization and monitoring. If previously the organization and monitoring of LAN was a priority just for large companies, in the modern world full of digital solutions, the organization and monitoring of data transmission networks is a factor in the success of any business. Considering that a huge amount of heterogeneous data is generated every minute in the network, the operational management of the LAN network infrastructure becomes a nontrivial task. Since the network infrastructure consists of a wide range of devices (switches, servers, routers, computers, etc.), and each of these devices utilizes a high number of application software, which is a source of immense amount of information, the forecast based on processing and analysis of this data would be best made by an algorithm of machine learning. The availability of a sufficiently accurate prediction of failures will be able to provide high reliability, safety and economic efficiency of the operation of equipment of segments of local area networks (hereinafter referred to as LAN).

SETTING OBJECTIVES

The main way to increase operational reliability is to predict malfunctions of devices. The forecast of device errors allows the identification of malfunctioning devices for their subsequent repair or decommissioning. As a result, the possibility of occurrence of malfunctions and equipment failures is minimized.

The operation of any equipment of the LAN segment involves the impact on it of a large number of different factors that causes changes in the technical condition, which eventually leads to failure. An essential feature of these factors is their random (stochastic) nature. The factors that have the most significant effect on the rate of change in the technical condition of the equipment include: ambient temperature, the presence of ventilation, the equipment manufacturer recommended air humidity and the degree of contamination of the circulating air inside the room where the equipment is located as well as temperature of the central processor unit and the availability of uninterrupted power supply. In addition to them, it is necessary to highlight such factors as: brand of equipment, traffic passing through the electric switchboard, schedule of equipment failures in the past, etc. The random nature of the considered factors leads to an unpredictable change in the technical condition of devices, their components, mechanisms, and therefore, its exploitation time until it fails. [1]

RESEARCH METHODS

To solve the problem of predicting the time of failures, various researchers proposed several models that differ in the set of input data, methods of its analysis and the form of presentation of the results:

- a method for predicting the time of equipment failure using the laws of the distribution of its resources obtained from repair statistics;
- a method for predicting failures based on statistical data on equipment (temperature, CPU load, equipment power failure)

The Holt-Winters' prediction algorithm was also considered.

The Holt-winters method is an improvement of the exponential

smoothing method of the time series. Exponential smoothing provides a clear view of the trend and allows you to make short-term forecasts.

The difference from exponential smoothing is the ability of the method to detect trends related to short periods at times immediately preceding the forecast ones, and to extrapolate these trends to the future. Although the method uses linear extrapolation, it is sufficient for most indicators of the current state of the local network.

The Holt-winters method is based on the fact that the time series under study can be represented as three components: the base component, the trend line, and the seasonal effect. The algorithm assumes that each of these components changes over time. Exponential smoothing is applied to the changing values of each component.

As in the exponential smoothing method, the forecast for the next period is calculated by applying coefficients α , β , and γ to the current forecast value . These coefficients are determined by the model parameters and can take values ranging from 0 to 1. At higher coefficient values, the past component values are taken into account more than the current ones, and at lower coefficients, the current component values have the greatest impact on the forecast.

The forecast is the sum of all three components:

$$\widehat{y_{t+1}} = a_1 + b_1 + c_{t+1-m} \quad (1)$$

where a_t is the base component; b_t is the trend line; c_t is the seasonal effect; m is the period of the season (cycle)

The new estimate of the base component is its current value, adjusted for the value of the seasonal coefficient. Since the new value of the base component depends on changes in the trend line, the trend forecast is added to the baseline coefficient:

$$a_1 = \alpha(y_1 - c_{t-m}) + (1-\alpha)(\alpha_t - 1 + b_{t-1}) \quad (2)$$

The new trend estimate is the difference between the new and old value of the base component:

$$b_1 = \beta(a_t - \alpha_{t-1}) + (1-\beta)b_{t-1} \quad (3)$$

The new estimate for a seasonal component is the difference between its current value and the corresponding base component:

$$c_1 = \gamma(\gamma_t - \alpha_t) + (1-\gamma)c_{t-m} \quad (4)$$

Formulas (2)–(4) are only used to get the current component values for a single time interval, since these stored values are recalculated in each iteration. At the first point in the series, the values of the base component and trend are not calculated, since there are no previous experimental values for calculating them. At the second point in the series, the smoothed value of the base component is assumed to be equal to its observed value, and the micro trend for this period is considered linear and calculated as the difference between the current and past response values.

Starting with the third point, it uses the formula (2) to(4): calculated on a smoothed value of the underlying components of the smoothed value and micro trend for last point of range and response for the current point, and then calculated new micro trend at its previous value and difference between the past and the only that an estimated smoothed value. Then the described procedure is repeated for all subsequent points in the time series.

Since the described method uses short-term forecasting, each time a value changes, a forecast of its next value is made, based on the history of previous measurements and the last value. You can use coefficients to vary the number of previous values of the measured value used for the forecast. Thus, it is possible to achieve the necessary level of detection of aberrations [6].

Each of these methods involves the use of its own forecasting procedure, which is determined by the type and nature of the source information. Each of these methods is performed in two stages. The first stage determines forecasting parameters and the second stage carries out the procedure for determining the operating time of the equipment until it fails. Thus, the forecasting accuracy of each of these methods directly depends on the choice of parameters.

Modern equipment is provided with the necessary number of sensors and control devices to track a large number of parameters of its operation. While forecasting errors, it is necessary to choose the most informative of them. To solve this problem, machine learning (ML) methods can be used [4].

Prediction of equipment failures based on ML methods has recently become an increasingly relevant task to be solved in such areas as transport systems, industry, biotechnology, and the field of IT services since the methods used are more effective than methods based on obscure logic, statistical methods etc. As an example, we can give suggestions in the field of predictive maintenance of equipment of such well-known companies as SAS, IBM [2], SAP [3], etc. However, the proposed approaches to constructing the forecast models are focused on learning on precedents, which in some cases cannot be applied in practice. For example, in situations where breakdowns are rare enough or there are not enough statistics. Therefore, the use of ML methods to improve the quality of forecasts of equipment failure in order to reduce the accident rate and costs is relevant.

The ML procedure requires a sufficient amount of data corresponding to the described mode of operation, since the collection of data corresponding to one or another failure requires observation for a long period of operation, and this data can be critically small, e.g. for new equipment. Therefore, training is carried out on data that correspond to the normal mode of functioning of the object of study (i.e. at a time when there were no breakdowns or other anomalies). Thus, the model learns to predict what the signal should be in normal operation. In the event that at a certain point in time the actual signal value differs from the predicted “normal” signal value, anomalous behavior is recorded and signals a possible breakdown.

RESEARCH RESULTS

Let’s consider the key stages of building a forecasting model with this approach. What is proposed is the following model, consisting of the stages of collecting parameters, training the model and forecasting. The first stage determines, initially, a list of all parameters that are collected from the environment of the equipment. These parameters include: ambient temperature, the presence of ventilation, the degree of contamination of the circulating air and the humidity recommended by the equipment manufacturer in the room where the equipment is located. Next, we collect data directly from the equipment

itself: the temperature of the central processor unit, the power supply of the equipment, the amount of data passing through the equipment. The positive effect of the redundancy of the collected parameters is that the forecasting model can take into account factors that may seem insignificant from the point of view of the impact on the failure. Moreover, it is important that the training sample consists only of data in which there are no failures and anomalies.

Let us consider the setup and operation of the algorithm as an example of forecasting network equipment failures, when the response time, temperature indicators of and traffic flowing through the equipment are used as input parameters. Data from the sensors was collected in a database during its operation for two days (Fig. 1).

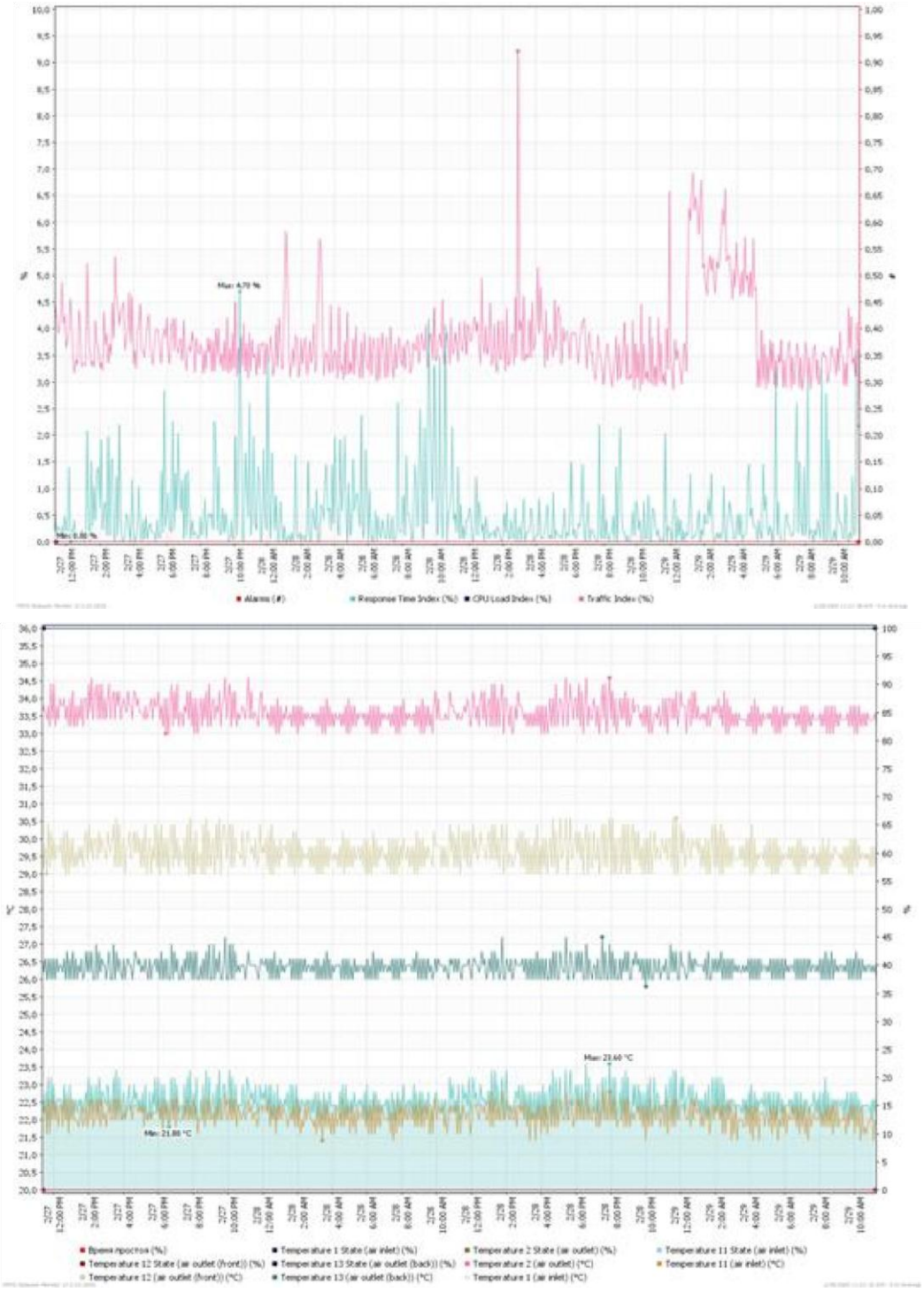


Figure 1. The graph of readings from the network equipment sensors

At the second stage, the forecasting model is trained. Random Forest was chosen as the ML. This algorithm is based on the construction of a large number of decision trees, each of which is built according to the set obtained from the initial training sample with return. [5] At the training stage, several models are built (in our case 3). Each model predicts the value of one of the parameters at the next point in time based on the previous values of all parameters. The input parameters for all models are the same. It is the value of all three signals for the previous 48 hours in increments of 1 hour (144 total values). The first model predicts the temperature of the processor of the network equipment, the second model predicts the index of equipment traffic, the third model predicts the index of the response time of network equipment at the next period of time (Fig. 2).

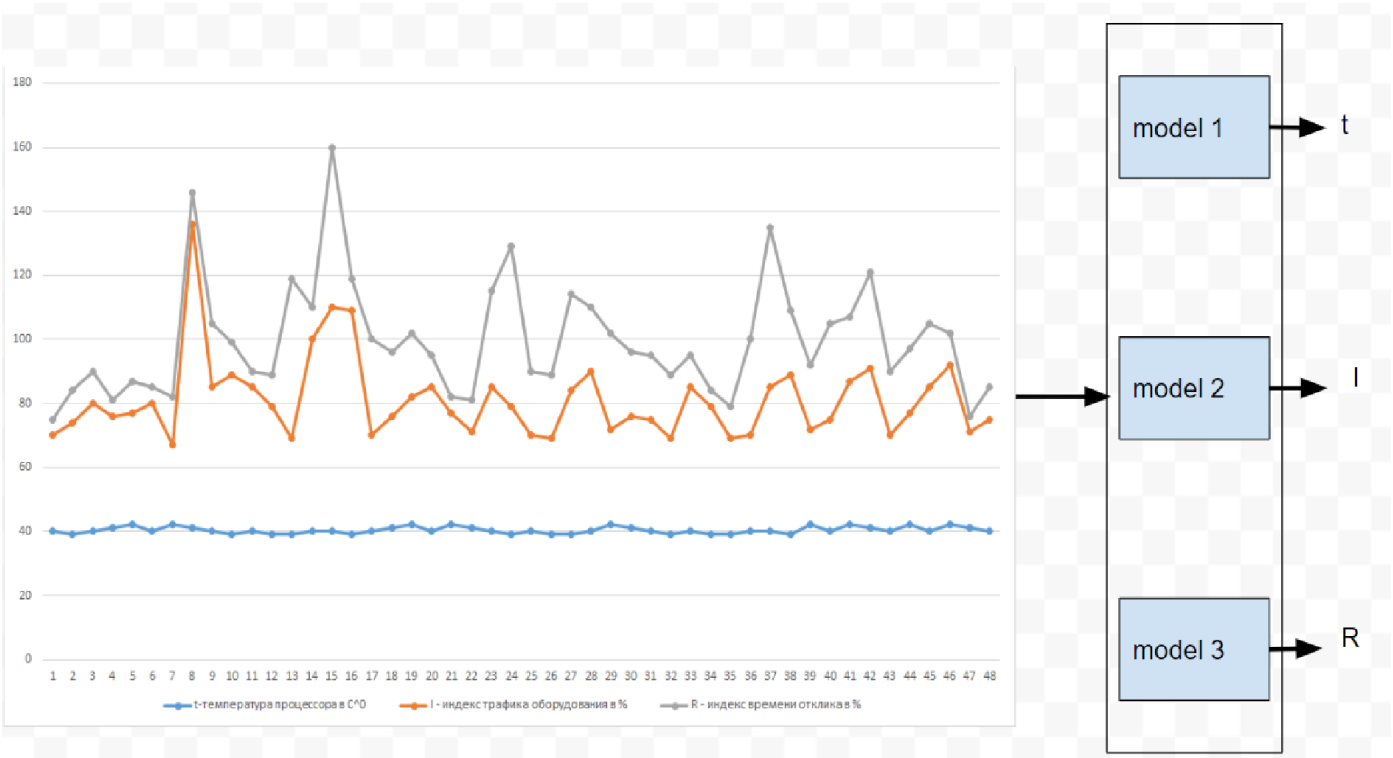
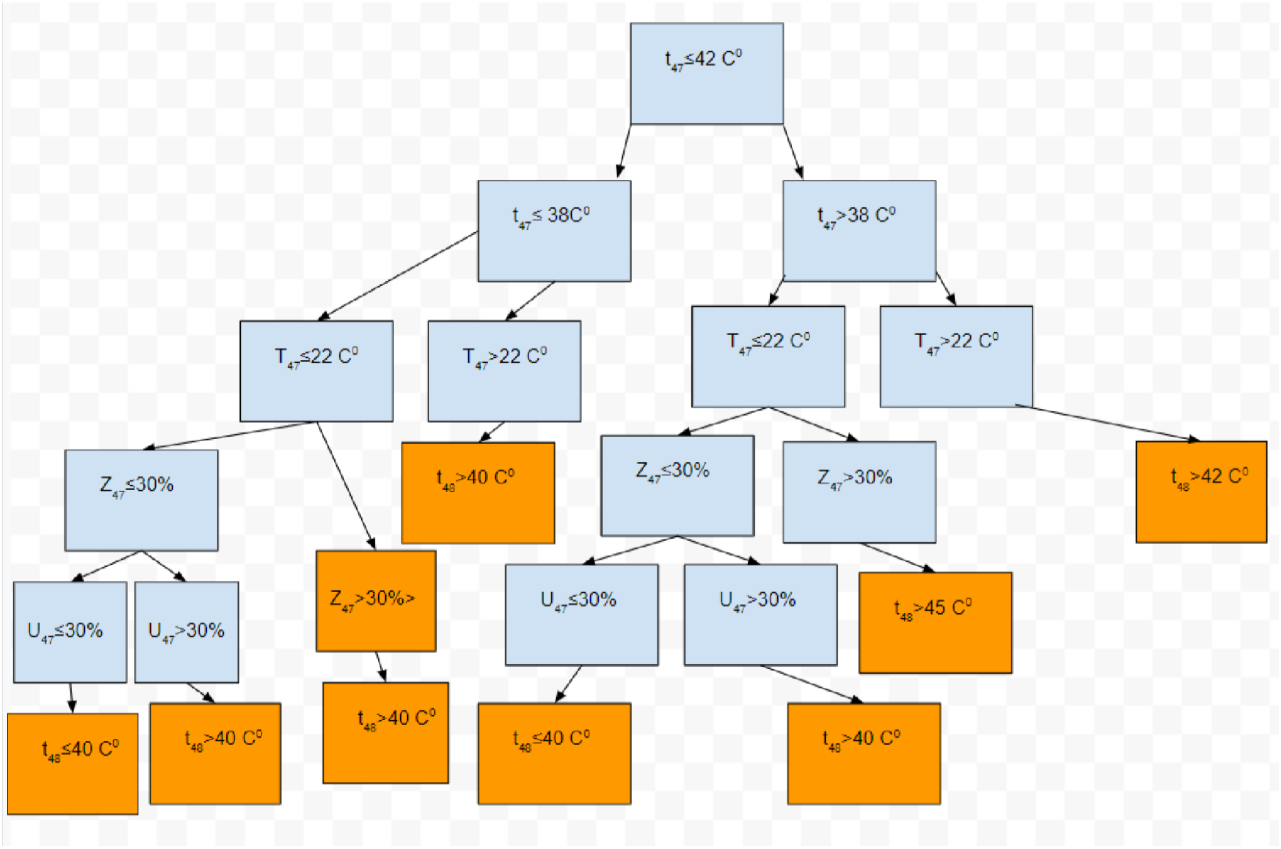


Figure 2. The structure of the forecasting model
"t" is the temperature of the processor of network equipment; "I" is the index of network equipment traffic;
"R" is the response index of network equipment

It should be noted that each of the models represents 100 binary decision trees. An example of a binary tree for determining the processor temperature of network equipment and the ambient temperature are presented in Figure 3.

In the third stage, after all the models are built, we must determine the algorithm according to which the system will determine the occurrence of an anomaly by three signals. For this, each of the three models is launched to predict further normal operation of the equipment. The latest values of sensors for the last 48 hours of operation are used as the starting data for forecasting. After receiving the next predicted signal value for the model, the input parameters at the first hour of equipment operation are discarded, and the predicted parameters for 49 hours are added to the input. Thus, each model is run to predict the values of the parameters during the next 2 days of operation (288 values for each of the three parameters) and to calculate the deviations of these predicted signals with the actual ones for the same period. As a result, we obtain data on deviations between the predicted and the actual signals at each moment in time. It is necessary to set a threshold for the maximum possible deviation of the forecast signal from the actual one to determine the possibility of equipment breakdown during this period of time. Moreover, the threshold must be set so as to prevent false alarms and minimize the omission of real breakdowns. If there is data on breakdowns, the threshold is set on the basis of these data, while also selecting the time interval during which the threshold is exceeded.



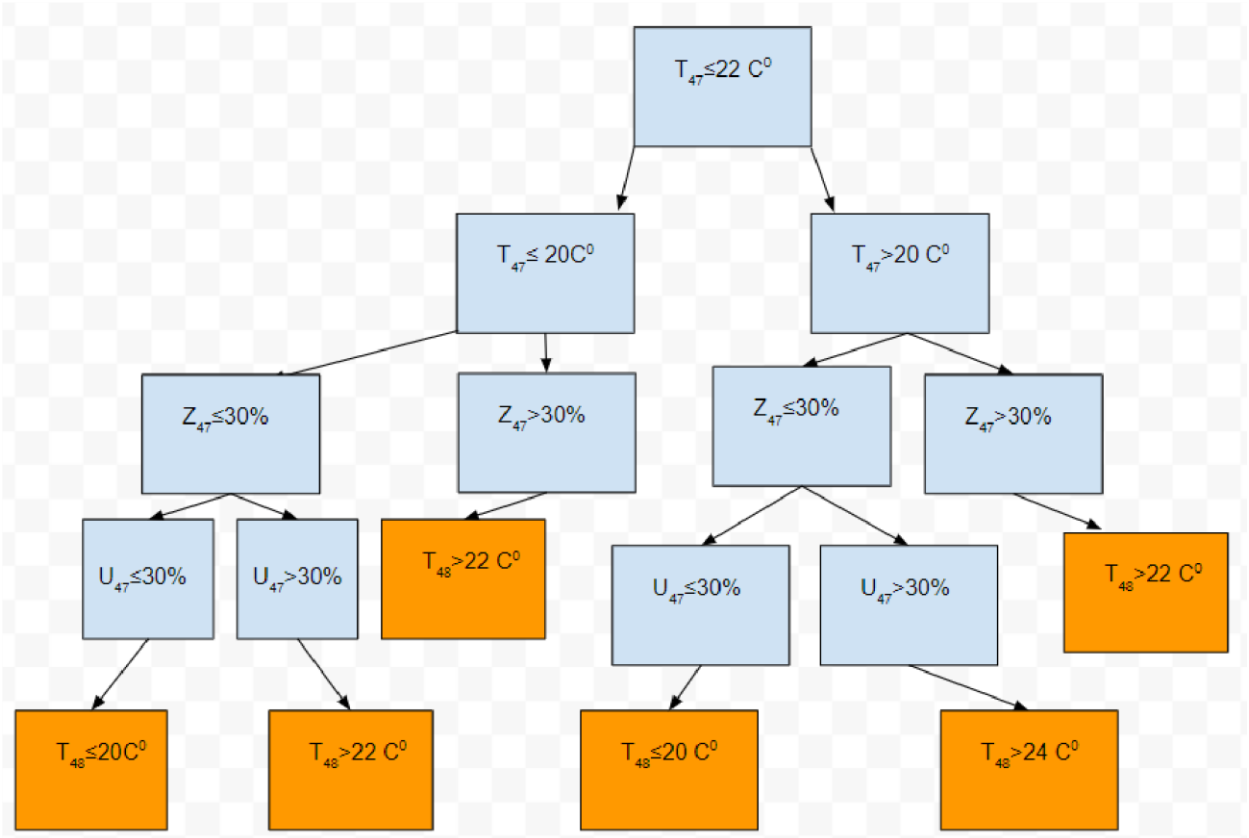


Figure 3. Decision trees for determining the temperature of the network equipment CPU (t) and the ambient temperature (T), where Z is the indicator of air pollution by dust, U is the sensors indicator of air humidity

CONCLUSION

Thus, in order to prevent network equipment failures, the article proposes a new approach to prediction of malfunction based on machine learning algorithms. This approach allows us to maximize the use of available information about the operation of equipment collected from its sensors and sensors around it while it functions normally. The model includes several sub-models predicting equipment failure, each of which uses the actual and the predicted indicators of the sensors. To forecast the signal, the Random Forest algorithm is used. For training, the only data used is the one during normal operation of the unit, in contrast to similar approaches in which training requires data before and during equipment malfunctions. The task of the machine learning algorithm is to predict the normal signal for each sensor using the signal values for the previous time interval. With the approach of predicting equipment failures in the context of a small number of breakdowns, as it is proposed in the paper, data on failures are used only when setting up the model to minimize false alerts about equipment errors and minimize omission of real breakdowns.

REFERENCES

[1] N. I. Shakhanov, I. A. Varfolomeev, E. V. Yershov, O. V. Yudina Forecasting of equipment failures in conditions of a small number of breakdowns. Bulletin of Cherepovets state University 2016 no. 6 P. 36-41-article

[2] Victor M. Predictive Analytics for efficient use of equipment. 2016. URL: https://filearchive.cnews.ru/files /reviews/2016_03_29/2_Maltsev.pdf -electronic source

[3] Oliver M. predictive maintenance and service (PdMS) - outline and value proposition. 2014. URL: https://blogs.saphana.com/wp-content/uploads/2014 /11/Predictive-Maintenance-Service-OutlineValue-Proposition.pdf - electronic source

[4] Alestra S., marki S., Burnaev E. V., Erofeev P. A., Papanov A., Bordry S., Silveira-Freixu S. a rare event of waiting and degradation of the trend for aircraft preventive maintenance / / 11th world Congress on computational mechanics, WCCM 2014, 5th European conference on computational mechanics, noise protection 2014 and 6th European conference on computational hydrodynamics, ECFD 2014 11, 2014. PP. 6571-6582. – article

[5] Chistyakov S. p. Random forests: an overview // Proceedings of the Karelian scientific center of the Russian Academy of Sciences, 2013, no. 2, PP. 117-136. – article

[6] S. Yu. Iskhakov, A. A. Shelupanov, S. V. Timchenko " Forecasting in the local network monitoring system", TUSUR Reports, no. 1 (25), part 2, June 2012. PP. 100-103-article