E3S Web of Conferences **549**, 08024 (2024)
TransSiberia 2024
https://doi.org/10.1051/e3sconf/202454908024

Using machine learning to forecast hard drive failures

Nikita Kodatsky[1], *Elena* Revyakina[1*], *Andrei* Gazizov[1], and *Vitaliy* Gunko[1]

[1]Don State Technical University, 1, Gagarin Sq., 344002 Rostov-on-Don, Russia

Abstract. The aim of this research is to develop an effective tool for assessing the condition of computer hardware based on an intelligent method. The outcome of this work not only facilitates the implementation of a data analysis tool into the realm of computer equipment servicing but also offers a practical example of its use for enhancing the reliability and efficiency of information systems. The findings of this research can be valuable for IT specialists responsible for technical support as well as for organizations seeking to optimize equipment servicing processes and improve their competitiveness in the market. Keywords: information technology equipment condition prediction, machine learning, SMART tests, equipment condition assessment model.

1 Introduction

One important aspect of maintaining the effective operation of information systems is the condition of computer hardware. The reliability and performance of computing equipment directly impact the functionality of business processes and the quality of services provided. In the rapidly evolving technological landscape and increasing complexity of computer hardware, the application of innovative methods for monitoring and predicting equipment status becomes necessary. One such method is the use of machine learning algorithms for analyzing and interpreting data obtained from SMART tests.

Data loss can be caused not only by «human errors, cyberattacks», but also by «equipment failures» [1], including computers and other storage devices, which are much more common even in normal mode [2,3], including computers and other storage devices, which are much more common even in normal operation [4]. One of the major factors of data loss is computer hardware failure [5]. And one of the main devices of a computer, concerning information storage, is its storage device, without which it is simply impossible to operate information. It is generally accepted that computer hardware is standardized devices that «must operate with a satisfactory depreciation factor of elements» and a minimum but «sufficient time to failure» [6].There are various factors that can lead to malfunction of computing equipment. In this regard, it is important to have a «hardware health monitoring system» [7] to prevent data loss and ensure the continuity of the computer system. One of the effective methods to solve this problem is to use intelligent methods to predict the failure of various devices. Current «intelligent methods for assessing

---

[*] Corresponding author: revyelena@yandex.ru

the state of computer hardware» that utilize «machine learning and artificial intelligence» [8], can help determine the likelihood of equipment failure before it occurs. In this way, preventive measures can be taken and faulty equipment can be replaced before it leads to data loss, making the research relevant.

Modern tools and approaches such as Google Coola cloud computing, including Python and machine learning libraries such as Scikit-learn, as well as other libraries for data handling and manipulation such as numpy and pandas will be used to address the objective. The paper will also utilize data and experimental results obtained on real computer systems.

The results of this paper represent a significant contribution to the field of information security and technology. The applicability of the work extends to various spheres of society, including business, science, medicine and other areas where computer technology is widely used [9]. The obtained results allow us to develop and implement our own domestic method of monitoring the state of information storage devices for enterprises in terms of performance not worse than foreign analogs. This approach to the problem of data security and control over information storage can replace the use of foreign analogs with closed source code. Thus, the research contributes to the strengthening of information security of organizations [10] and makes it possible to have more reliable tools to control information resources in the context of complex political circumstances and geopolitical relations with the sphere of products and developments.

2 Materials and methods

2.1 Manual analysis of the data

Hard disk drives (HDDs) are one of the most common storage devices in computers and servers. They play a key role in storing, accessing, and processing information. Therefore, understanding the condition and performance of hard disks is essential for the effective functioning of computer systems. Various statistics are used to assess the health and performance of hard disks, including SMART (Self-Monitoring, Analysis, and Reporting Technology) metrics. SMART is a technology that «allows you to monitor and analyze the performance of a disk device, warn of possible problems and predict failures» [11]. Studying SMART statistics is an important aspect for effective management and support of hard disk drives. Understanding what these metrics tell us about the health of the disk «allows us to take timely action to prevent data loss and reduce the risk of system failure» [12]. In addition, analyzing SMART statistics is also significant in the field of information security. Determining the state of a hard disk using SMART metrics can help identify potential problems related to data integrity or possible security breaches.

Thus, the use of SMART statistics not only contributes to effective hard disk management, but also plays an important role in ensuring information security and protecting data from potential threats.

Each drive includes Self-Monitoring, Analysis and Reporting Technology (SMART), which provides internal information about the drive. SMART is a monitoring system built into hard disk drives that reports on various attributes of a given disk's health. More than 70 SMART statistics are available, but only five will be sufficient to know for the task of predicting WD failure. The attributes are described in Table 1.

Table 1. SMART test attributes and its description

Attribute	**Description**
CMAPT 5	Number ofsectorsreallocated
CMAPT 187	Reportedincorrigibleerrors
CMAPT 188	Command interruption
CMAPT 197	Current number of waiting sectors
CMAPT 198	Irreparablenum berof sectors

When the value of one of these attributes is greater than zero, there is reason to pay attention. It is also «necessary to monitor RAID status to identify potential disk problems». These tools usually only report exceptions, so the number of investigations can be monitored at any time, even though the number of HDDs is significant.

According to the statistics provided by Backblaze Vault [13], by analyzing the state of the HDDs (see Table 2), we can say that the percentage of working HDDs with one or more of the five SMART indicators greater than zero was 4.2%, when the faulty disks with the same indicators greater than zero was 76.7%.

Table 2. Percentage of faulty and working WDs by selected attributes

HDD state	SMART 5	SMART 187	SMART 188	SMART 197	SMART 198
Functional	1,1%	0,5 %	4,8 %	0,7 %	0,3 %
Faulty	42,2 %	43,5%	44,8%	43,1 %	33 %

The presence of this indicator with a value greater than zero may mean nothing at this point. For example, a drive may have a raw SMART 5 value of two, which means that two sectors of the drive have been reassigned. By itself, such a value means little until it is combined with other factors. The evaluation process «may require a fair amount of intelligence» (both human and artificial) to come to the conclusion «that the operating engine will fail». The only way this helps is if multiple SMART errors are observed.

Next, consider the correlation between the observed attributes in Table 3 [14]. The table reflects the relationship between the attributes under consideration, which helps in determining which specific SMART parameters are more related to each other, as well as identifying contradictions or dependencies between different attributes.

Table 3. Correlation between SMART attribute statistics

	SMART 5	SMART 187	SMART 188	SMART 197	SMART 198
SMART 5	1	0,034	0,026	0,064	0,043
SMART 187	0,034	1	0,007	0,025	0,033
SMART 188	0,026	0,007	1	0	0,006
SMART 197	0,064	0,025	0	1	0,808
SMART 198	0,043	0,033	0,006	0,808	1

In most cases, the statistics have little correlation and can be considered independent. Only SMART 197 and 198 have a good correlation, i.e. they can be considered as one indicator instead of two. But it is advisable to keep collecting these statistics together for two reasons: 1) «the correlation is not perfect», so there is room for error and 2) «not all drive manufacturers report both attributes» [15]. Understanding the correlation or lack thereof can help in deciding what to do with a WD. For example, a drive reported a raw SMART 5 value of 10 and a raw SMART 197 value of 20. The assumption is that the HDD is wearing out and should be replaced. Considering that if the same drive had a raw

SMART 197 value of 5 and a raw SMART 198 value of 20 and no other errors, one could delay replacing the drive while waiting for additional data such as error rates. The SMART statistics above, with the exception of SMART 197, are cumulative in nature, meaning that the time period over which errors were recorded should be considered, not just the number of non-zero values.

Proper use of SMART statistics requires systematic monitoring and analysis of the metrics. By regularly monitoring disk health and applying SMART statistics guidelines, you can improve system reliability and ensure data storage security. In conclusion, understanding the meaning and application of SMART statistics is an integral part of hard disk drive management and support. Analyzing SMART statistics can help prevent data loss, alert you to potential problems, and improve system security. Utilizing this information allows you to take timely action and ensure that your computer systems are operating reliably in a rapidly changing information environment.

2.2 Data for model fit

One of the effective methods of solving the problem under study is the use of intelligent methods of predicting the failure of various devices. The purpose of this section is to substantiate the effectiveness of an intelligent method for assessing the state of computer hardware using the Random Forest algorithm and bagging and boosting approaches to prevent data loss on computer information storage devices.

- The source of data is the device operating system (Windows or Linux), namely the following system sources: SMART attribute monitoring technologies (smartmontools or CrystalDiskInfo), operating system event log and command line utilities. The 5 attributes of SMART-statistics data are of the greatest interest for determining hard disk failure, namely [16]:
- SMAPT 5 – number of reallocated sectors;
- SMAPT 187 – faultmessages;
- SMAPT 188 – time required to wait for commands;
- SMAPT 197 – current number of sectors in standby mode;
- SMAPT 198 – number of uncorrected sectors.

To train the model more correctly, it was decided to use the collected raw hard drive test data for the years 2022 - 2023 from Backblaze using this data in their 67814 hard drive processing centers. [9]. The company uses Smartmontools [5] to collect the SMART statistics data. The collection is done once a day for each hard disk. Thus, «several elements» such as disk model, serial number, etc. are added and «a line in the daily log for each disk is created» [10]. Disks that have failed are marked as such and their data is no longer logged. Sometimes «disks are removed from service» even if it has not failed, such as when a company «upgrades a Storage Pod» by replacing 1TB disks with 4TB disks . In this case, the 1TB disk is not marked as failed, but the SMART data is no longer recorded.

Every day, the Backblaze data center takes a snapshot of every running hard drive in its research center. This «dataset includes basic information about the disk» as well as its SMART statistics . A daily snapshot of a single disk is «one record or row of data». All disk snapshots for a particular day are collected into a file consisting of rows for each active hard disk. This file is in the format «*.csv» (comma separated values). Thus, each day a file with the name in the format YYYYY-MM-DD.csv is formed, for example, 2023-07-01.csv and has the following structure [8]:

1)the first line of each file contains the column names, the remaining lines are historical data of the drive status.

2) columns contain the following information:

- date - file date in yyyyy-mm-dd format;

- serial number - the drive serial number assigned by the manufacturer;
- model - the drive model number assigned by the manufacturer;
- capacity - the capacity of the drive in bytes;
- failure - contains «0» if the drive is OK. Contains a «1» if this is the last day the drive was operational before the failure.

2.3 Choosing anintelligent method

Different machine learning algorithms can be used for the task of predicting hard disk drive failure based on SMART marked-up data. Each model has its advantages and disadvantages, and the choice of a particular model depends on the specifics of the problem, data availability, accuracy and interpretability requirements. I have highlighted some of the suitable ones:

1. Random Forest. Can achieve high accuracy and robustness to overfitting due to ensemble of decision trees. Can process large amounts of data with multiple attributes. It is possible to estimate the importance of attributes, which is useful for understanding which SMART attributes have the most impact on failure. But such models on real-world and global tasks can be quite complex and require significant computational resources for training and predictions. And its interpretability may be lower than some other approaches.
2. Gradient Boosting. One of the most powerful and widely used techniques, which often shows high accuracy in classification and regression tasks. The method builds a model sequentially, which allows it to effectively correct errors of previous models. Also, requires careful tuning of hy perparameters to achieve optimal performance. May be more prone to overfitting than random forest when insufficient data are available.
3. Logistic regression. Easy to interpret results. Has fast learning and prediction speed compared to more complex models. Well suited for binary classification tasks. However, assumes a linear relationship between the features and the target variable, which may not always be true for SMART data.
4. Support Vector Method (SVM). Effective in high-dimensional spaces and in the presence of clear class separation. Possess verisimilitude (ability to process a wide range of data) due to different kernels for processing nonlinear dependencies. But require intensive computations for large amounts of data, and kernel selection and tuning can significantly affect model performance.

I have chosen the Random Forest model to predict hard disk failure based on the labeled SMART test data. In my opinion, it is a powerful algorithm that combines the advantages of multiple decision trees to improve the accuracy and robustness of the model. Random Forest handles large datasets well and can automatically take into account the importance of features, which makes it particularly suitable for analyzing SMART data.

Random Forest is an ensemble method that builds multiple decision trees during training and produces an average prediction for classification or regression. It achieves high prediction accuracy while reducing the risk of overtraining due to randomness mechanisms in the selection of features and samples to build the trees. SMART test data often includes many different attributes reflecting the state of hard disks. A random forest can efficiently process such datasets by automatically identifying the most relevant attributes for failure prediction. One of the basic principles of such an algorithm is the use of subsamples of attributes for each tree, which reduces the influence of irrelevant or weakly influential attributes on the target variable and increases the overall accuracy of the model. Although the random forest model itself may seem less interpretable compared to a single decision tree, it provides useful information about the importance of attributes. Understanding which SMART attributes most influence predictions may be important for further analysis and understanding of the causes of hard disk drive failures.

The effective use of the Random Forest algorithm with bagging and boosting approaches provides «significant opportunities for predicting» the failure of computer hardware and preventing potential data loss of computer hardware failure and prevent potential data loss. This combined approach combines the advantages of bagging (averaging the results of multiple models) and boosting (adaptive weighting of model errors) to promote robust and accurate models. This approach builds a strong classifier that is «capable of adapting to different data and conditions» [17], improving the accuracy of predictions. Bagging and boosting approaches improve the generalizability of the model, allowing it to efficiently handle complex data and respond quickly to changes in the state of the equipment. The advantages of the algorithm include :

- has high prediction accuracy (for most problems it performs better than linear algorithms, and the accuracy is comparable to boosting);
- has almost no sensitivity to outliers in the data due to random sampling by bootstrap method;
- is not sensitive to scaling (any monotonic transformations) of feature values, due to random subspace sampling;
- does not require careful parameter tuning, works well «out of the box»;
- is able to efficiently process data with a large number of features and classes;
- is rarely retrained, in practice adding trees almost always only improves the composition (up to a certain limit);
- handles missing data well, retains good accuracy if most of the data is missing;
- can be extended to unmapped data, leading to the ability to do clustering and data visualization, outlier detection;
- can be easily parallelized and scaled (increase the number of trees and their depth).

The drawbacks of the algorithm include :

- unlike a single tree, the results of a random forest are difficult to interpret;
- algorithm performs worse than many linear methods when there are a lot of sparse features in the sample (texts, bag of words);
- does not know how to extrapolate data, unlike linear regression;
- is prone to overtraining on some tasks where the data is not very noisy;
- for data including categorical variables with different number of levels, the random forest is biased in favor of features with more levels: when a feature has many levels, the tree will adjust more strongly to these features, since a higher value of the optimizing functional (information gain) can be obtained on them;
- large size of the resulting models, which requires O(N·K) memory to store the models, where K is the number of trees.

Thus, the choice of random forest for predicting hard disk failure on the labeled SMART test dataset is due to its ability to handle large and complex datasets, its high accuracy and robustness to overfitting, and its ability to interpret feature importance. These qualities make random forest particularly suitable for tasks where reliable and accurate prediction based on a large number of features is required, as is the case with hard disk drive SMART test data. The combination of such methods not only provides high accuracy in predicting hardware failures, but also allows for rapid response to any changes or malfunctions, minimizing the likelihood of serious failures. This approach ensures the stability of computer systems and provides a high level of data security, which is a critical aspect in today's information technology and business environment. Next, let us proceed to the implementation of the intelligent method model itself in the task of predicting the failure of a computer information storage device.

2.4 Using the Random Forest algorithm in the problem of predicting the failure of a computer's information storage device

Let us display a matrix of scatter diagrams (see Figure 1) to establish the dependencies between the 5 SMART parameters of the dataset used. Each diagonal of the matrix shows the distribution of one parameter, and the off-diagonal shows the scatter plots for each pair of parameters. Note that, as described earlier, the 5-attribute data are closely related and correlated.

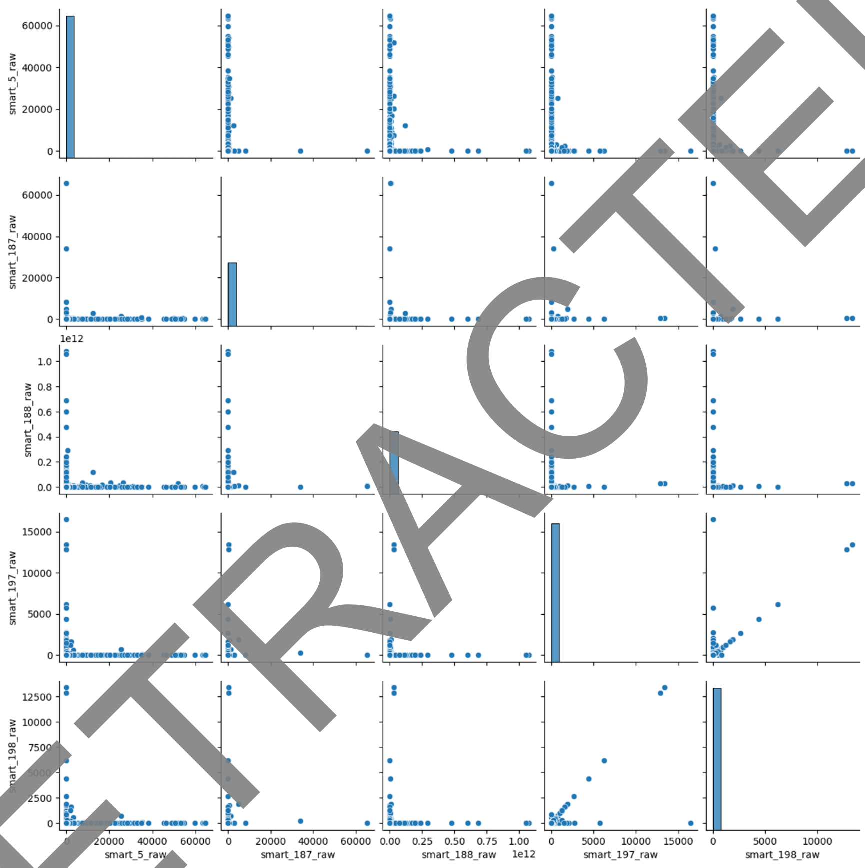

Fig. 1. Scatter matrix of 5 SMART attributes of the training sample

– It is important to note that for training, only values with high correlation on failure and the class labels labeled for them will remain from the dataset :

SMART 5 - number of sectors reallocated;
SMART 187 - reported uncorrectable errors;
SMART 188 - command timeout;
SMART 197 - number of current pending sectors;
SMART 198 - number of bad sectors;
Diskstatusclasslabel (0 - failed, 1 – functionin properly).

Based on the results of model training, it is important to evaluate its qualitative indicators. Let's consider 5 metrics to evaluate the correctness of the decisions made by the model and its bias to the data. Each of these metrics provides information about different aspects of model performance, namely:

- Accuracy - the proportion of correctly predicted classes relative to all examples;
- Precision - the proportion of true positive examples among all positive predictions;
- completeness (Recall) - the proportion of true positive examples that were predicted correctly;
- F1-measure (F1-Score) - the harmonic mean of accuracy and completeness;
-Matthews correlation coefficient - a measure of binary classification quality that takes into account false positive and false negative predictions.

We also display the error matrix (confusion matrix) to evaluate the performance of the classification model (Fig. 2). The results of the metrics obtained during the study are as follows:

```
[ Model ]>>>Random Forest classifier
[ The accuracy ]>>> 0.9998749937496875
[ The precision ]>>> 0.9997512809033477
[ The recall ]>>> 1.0
[ The F1-Score ]>>> 0.9998756249844531
[ The Matthews correlation coefficient ]>>>
0.9997500123101655
```

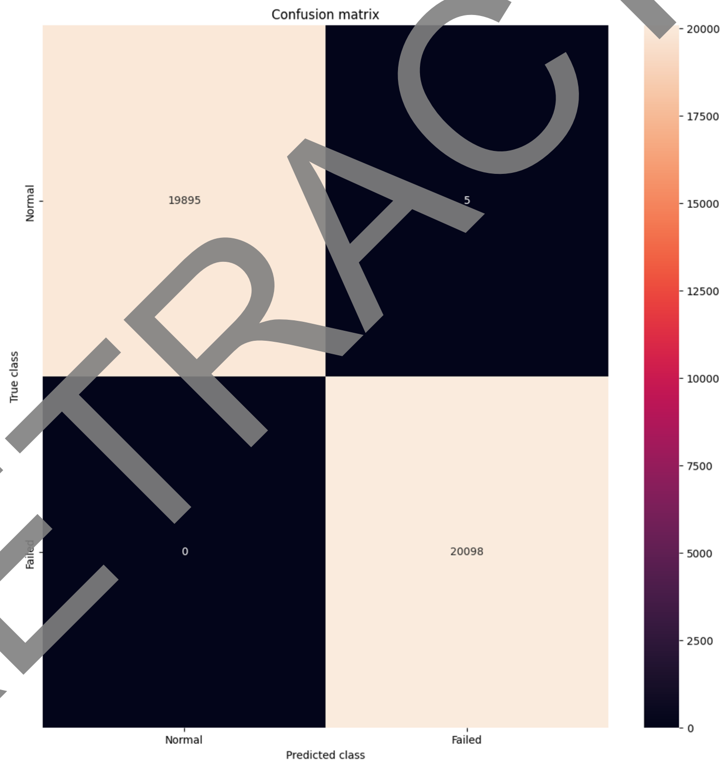

Fig. 2. Model error matrix

Then we will formalize the obtained results and make a conclusion about the work done.

3 Results and discussion

3.1 Analysis of results and comparison with expectations

From the results obtained, the following conclusions can be drawn about the performance of the trained Random Forest model:

1. Accuracy. With a high accuracy of about 99.99%, the model correctly classifies examples from the test dataset. This means that the model correctly predicts the class of an example in almost all cases.
2. Precision. Precision is also very high at about 99.98%. This indicates that of all the examples that the model predicted as positive, almost all are indeed positive.
3. Recall. Recall is 1.0, which means that the model correctly classifies all true positive cases from the test dataset. This is a good score and indicates that the model effectively detects all positive cases.
4. F1-measure. The F1-measure, which is the harmonic mean between accuracy and completeness, is also very high at about 99.99%. This indicates a balance between the accuracy and completeness of the model.
5. Matthews Correlation Coefficient. The MCC is also close to one, indicating a very strong correlation between the model predictions and the actual class values.

Overall, these results indicate that the Random Forest model is well trained and shows high performance on the test dataset .

Considering an error matrix where the rows represent the actual classes and the columns represent the predicted classes. Here is an interpretation of the results:

1. True Positives (TP), the upper left cell (19895) corresponds to the situation where the model correctly classified normal (negative) examples as normal.
2. False Positives (FP), upper right cell (5) corresponds to a situation where the model incorrectly classified normal examples as denied (false positive).
3. False Negatives (FN), lower left cell (0) corresponds to a situation where the model misclassified failed examples as normal (false negatives).
4. True Negatives (TN), the lower right cell (20098) corresponds to the situation where the model correctly classified the failed examples as failed.

Based on the matrix in Figure 10, the total number of correctly classified examples (True Normal and True Negatives) is 39893, and the number of incorrectly classified examples (False Normal and False Negatives) is 5.

Also, visualizations of tree ensemble elements are of interest. I demonstrate the obtained 1st, 2nd, 3rd and 50th constructed decision trees in Figure 3 - 6.

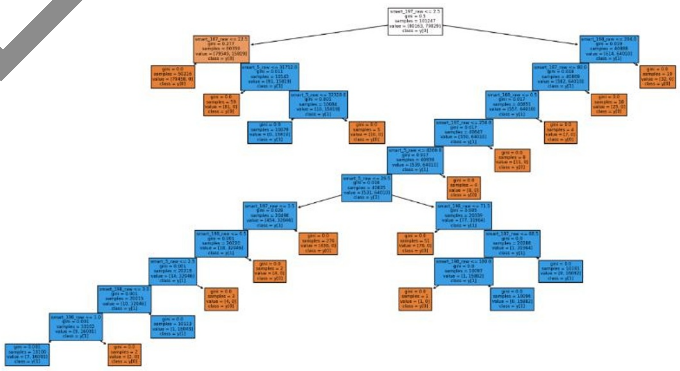

Fig. 3. The first decisive tree

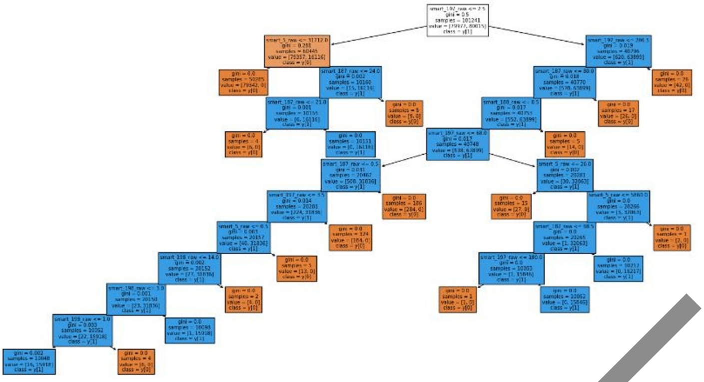

Fig. 4. The second decision tree

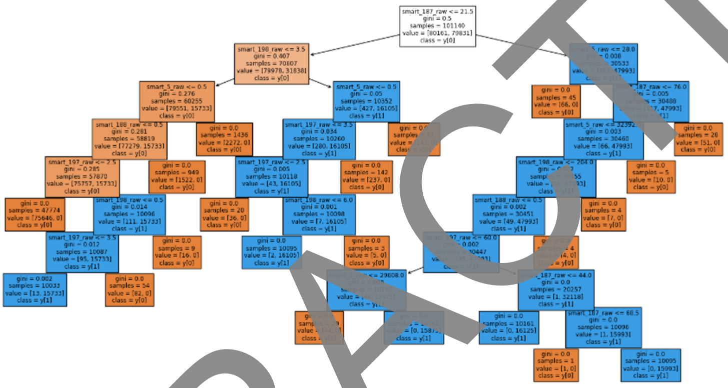

Fig. 5. The third decisive tree

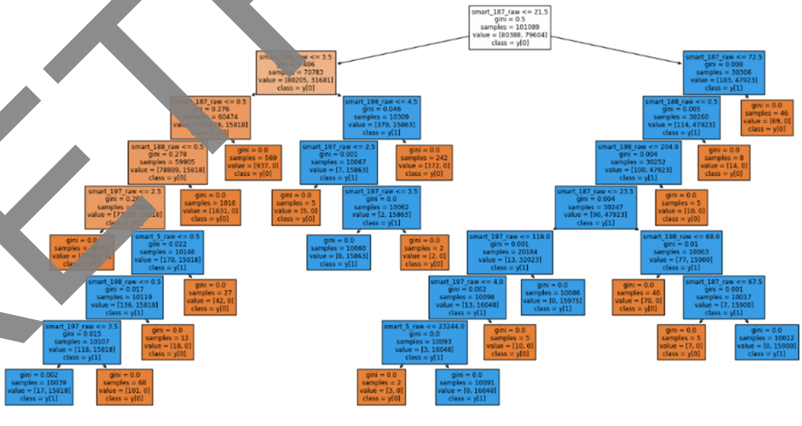

Fig. 6. The fiftieth decisive tree

The random forest method has a complex formalization, which is its main disadvantage. But evaluating the picture in general, we can see that trees undergo changes and try to identify the most correlated attributes for prediction. In this problem, the exact explanation of the solution proposed by the model is not the key aspect, it is more important to learn about errors in advance and try to identify the disk problem, which the model copes with very well [18,19].

4 Conclusion

Thus, this paper developed a computer hardware state estimation tool based on Random Forest algorithm using historical SMART test data. The results of the work showed that the developed model has high accuracy and ability to correctly classify both normal and failed components. Confusion matrix analysis confirmed that the model demonstrates good ability to correctly identify both classes. The model has a high ability to correctly identify normal states (True Negatives), which is good. The model also has a certain number of False Positives, which can lead to unnecessary alarms or unnecessary equipment checks. The model always manages to accurately predict failures (False Negatives), which eliminates missing real problems.

It is important to note that the training took place under severe computational power constraints. The data set was taken only for two days of observations of SMART-tests in 2023, because the amount of RAM did not allow to operate a large set, and in the experiment was taken the maximum of the provided virtual resources of the computing machine (Fig. 7).

Fig. 7. Computational resources used during training

In general, the results of the work indicate the prospect of using the Random Forest algorithm to create a system for monitoring the state of computer equipment. Further research and improvements of the model can help to even more accurately and reliably identify failed components and improve the efficiency of equipment maintenance, providing results no worse than analog systems, but using domestic proprietary development, which can be integrated and used for the needs of each individual enterprise.

References

1. National Standard of the Russian Federation GOST R ISO 13381-1-2011 Condition monitoring and diagnostics of machines. Prediction of technical condition. Part 1. General guidance. Date of introduction 2012-12-01. Prepared by Autonomous Non-Commercial Organization "Research Center for Control and Diagnostics of Technical Systems" (ANO SIC CD). Introduced by the Technical Committee for Standardization TC 183 "Vibration, Shocks and Technical Condition Control". Approved and put into effect by the Order of the Federal Agency for Technical Regulation and Metrology dated November 16, 2011 № 553-st.
2. Shklyar, V. N. Reliability of control systems: textbook (Tomsk: Tomsk Polytechnic University, 2009)
3. D. Anderson, J. Dykes, E. Riedel, *More than an interface - scsi vs. ata,* in Proceedings of the 2nd USENIX Conference on File and Storage Technologies (FAST'03), 2003, pp. 245 – 257.
4. Chandola, V., Banerjee, A., Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), **41(3),** 15.

5. Smartmontools official website - URL http://www.smartmontools.org/ (accessed 20.09.2023) - Text electronic.
6. Backblaze official website. Article "Backblaze Vaults: Zettabyte-Scale Cloud Storage Architecture." - URL https://www.backblaze.com/blog/vault-cloud-storage-architecture/ (date of reference: 20.09.2023) - Text electronic.
7. Official website of Backblaze. Article "Hard Drive SMART Stats." - URL https://www.backblaze.com/blog/hard-drive-smart-stats/ (date of reference: 22.09.2023) - Text electronic.
8. M. M. Botezatu, I. Giurgiu, J. Bogojeska, D. Wiesmann, *Predicting Disk Replacement towards Reliable Data Centers,* in: Proceedings of the 22nd ACM SIGKDD International Conference on 68 Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 2016, pp. 39–48, doi: 10.1145/2939672.2939699
9. Backblaze official website. Article "What SMART Stats Tell Us About Hard Drives." - URL https://www.backblaze.com/blog/what-smart-stats-indicate-hard-drive-failures/ (accessed 21.09.2023) - Text electronic.
10. N. Kolokas, T. Vafeiadis, D. Ioannidis and D. Tzovaras, *Forecasting faults of industrial equipment using machine learning classifiers,* 2018 Innovations in Intelligent Systems and Applications (INISTA), 2018, pp. 1-6, doi: 10.1109/INISTA.2018.8466309.
11. Russian Federation. Laws. On personal data: Federal Law of the Russian Federation No. 152-FZ. [adopted by the State Duma on July 8, 2006: approved by the Federation Council on July 14, 2006]. - Moscow: Kremlin: Codex, 2021. - 24 c. // ConsultantPlus.
12. Russian Federation. Laws. About information, information technologies and about protection of information: the Federal law of the Russian Federation No. 149-FZ: the text with amendments and additions for March 20, 2021: [adopted by the State Duma on July 8, 2006: approved by the Federation Council on July 14, 2006]. - Moscow: Kremlin: Codex, 2021. - 24 c. // ConsultantPlus.
13. Breiman, L. Random forests. Machine Learning. New York: Springer, 2001, 32 p. (date of reference: 18.03.2024) - Text electronic.
14. Liaw, A., Wiener, M. Classification and regression by random forests. London: R Foundation, 2002, 22 p. (date of reference: 18.03.2024) - Text electronic.
15. Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., Lawler, J. J. Random forests for classification in ecology. New York: Springer, 2007, 582 p. (date of reference: 18.03.2024) - Text electronic.
16. Breiman, L. (2001). Random forests, Machine learning **45(1):** 5–32.
17. Kazaryan M. Analysis of the Possibilities of Carrying Out Attacks on the Functions of Transferring Control to Operating System Console Using Active Intelligence Methods International Research Journal of Multidisciplinary Scope, 2024, **5(2),** p. 516–534
18. Korochentsev D. A. et al. Import-substituting technologies for information security and data protection. - 2021.
19. Gazizov A. R., Gazizov E. R.// Young Researcher of the Don. 2018. **1 (10).** 6-9.