Mechanical Equipment Failure Prediction Based on Machine Learning

Zihao Zheng [1, *], Chong Li [1], Huimin Zhou [2]

[1] School of Communication Engineering, Tongda College of Nanjing University of Posts and Telecommunications, yangzhou, China, 225127

[2] School of Mathematics and Big Data, Chongqing University of Arts and Sciences, chongqing, China, 402160

* Corresponding Author Email: 2514228334@qq.com

Abstract. Mechanical equipment plays a central role in the manufacturing industry. However, they often suffer from problems such as aging and wear, which pose a major challenge to maintaining the stability of the production environment and improving economic efficiency. In order to solve this problem, this article conducted an in-depth study using the usage and failure data of a certain enterprise's mechanical equipment. First, this article establishes an XGBoost classification model to determine whether mechanical equipment has malfunctioned. The model performed well, with performance metrics showing high precision, recall, and F1 scores. Furthermore, this paper used the XGBoost model to successfully predict the faults of 19 mechanical equipment, classified these devices, and accurately determined the types of faults, including heat dissipation faults, power faults and overload faults. Through the weight analysis of various types of fault data using the CRITIC weight analysis method, the main causes of various types of faults are clarified, which provides an important reference for further improvement and maintenance of this mechanical equipment.

Keywords: XGBoost Classification Model, Classification Prediction, CRITIC Method.

1. Introduction

"Chang'e" captures the moon, "Jiaolong" enters the sea, "Mozi" transmits messages, and "Zhurong" detects fire... In the past ten years of the new era, my country's basic research and original innovation have been continuously strengthened, and core key technologies have made breakthroughs. The 20th National Congress report also pointed out that our country must accelerate the implementation of an innovation-driven development strategy and accelerate the achievement of high-level scientific and technological self-reliance and self-reliance. Among them, manufacturing is the main body of the national economy, and machinery and equipment are the core of manufacturing. In industrial production, mechanical equipment will cause problems such as aging and wear during operation, which will lead to the occurrence of various failures and affect production. Based on the usage of mechanical equipment, predicting potential failure risks in advance so as to conduct precise maintenance and repairs is crucial to improving the stability of the production environment and improving corporate efficiency.

A large number of scholars have done relevant research work on fault prediction of advanced mechanical equipment. Fan Guodong et al. [1] built an industrial machinery equipment fault prediction model based on the Production provided technical support. Tian Hui et al. [2] proposed using the single classification algorithm SVDD to train the model to achieve equipment failure prediction. The characteristic of this method is that only normal samples are used to train the model without the need for a large number of fault samples. Huang Chen [3] combined the ensemble learning method to design a Stacking-based multi-model fusion prediction method for shield faults and used the Bayesian optimization algorithm to optimize the model. Experiments have proven that the method has higher accuracy. It can more accurately predict possible failures during the operation of the shield machine. Liang Chen [4] uses artificial intelligence methods to easily achieve accurate fault prediction without the need for professional time-frequency analysis knowledge. This is of great significance to improving equipment reliability and reducing costs. Zhang Jian [5] was able to synthesize more fault data through a generative adversarial network, thereby improving the performance of the model, especially when there are insufficient samples in the minority category. Wang Chunzhu [6] is improving the real-time prediction method of equipment faults through deep learning technology, including data preprocessing, feature extraction and real-time update model to improve the accuracy and real-time performance of fault prediction. Huang Pengcheng [7] is using multi-task network and attention mechanism, adaptive cost-sensitive matrix multi-task learning algorithm and other technologies to solve the data balance and imbalance problems in hydraulic system fault prediction to improve the accuracy and practicality of prediction.

Based on the above research, the XGBoost classification model established in this article performs well, with high precision, recall and F1 score, and is used to determine whether mechanical equipment is faulty. Finally, the XGBoost model was successfully used to predict and classify the faults of 19 mechanical equipment, including heat dissipation faults, power faults, and overload faults. Through weight analysis of fault categories, this article identifies the main causes of various faults and provides key fault diagnosis information for enterprises. The theoretical and practical significance of this research is to help enterprises effectively diagnose mechanical equipment faults and use the CRITIC weighting method to deeply analyze the causes of faults. It provides valuable information for improving and maintaining production equipment and is expected to have a positive impact in the manufacturing industry.

2. XGBoost discriminant prediction model

In the classification model of machine learning, the XGBoost algorithm uses a step-by-step forward additive model. Through integrated learning, the Boost method is used in series to correlate the upper and lower models and build a tree through the correlation of residuals. Since XGBoost has the advantages of parallel optimization, custom loss, higher efficiency and prediction accuracy, the XGBoost discriminant prediction model is established to predict whether mechanical equipment will malfunction.

2.1. XGBoost discriminant prediction model steps

Step 1: Determine the training set and establish the XGBoost classification model.

Step 2: Calculate feature importance.

Step 3: Apply the established XGBoost model to the training set, verification set, and test set to evaluate the effect of the XGBoost discriminant prediction model.

Step 4: Since the XGBoost classification model is random and the calculation results are different each time, in order to facilitate subsequent prediction use, the XGBoost discriminant prediction model with a good training effect needs to be saved, and subsequent substitution data can be used directly.

2.2. Brief description of XGBoost discriminant prediction model

Since the derivation process of the XGBoost algorithm is complicated, the author quotes a picture from [8], as shown in Figure 1, to illustrate the specific process.

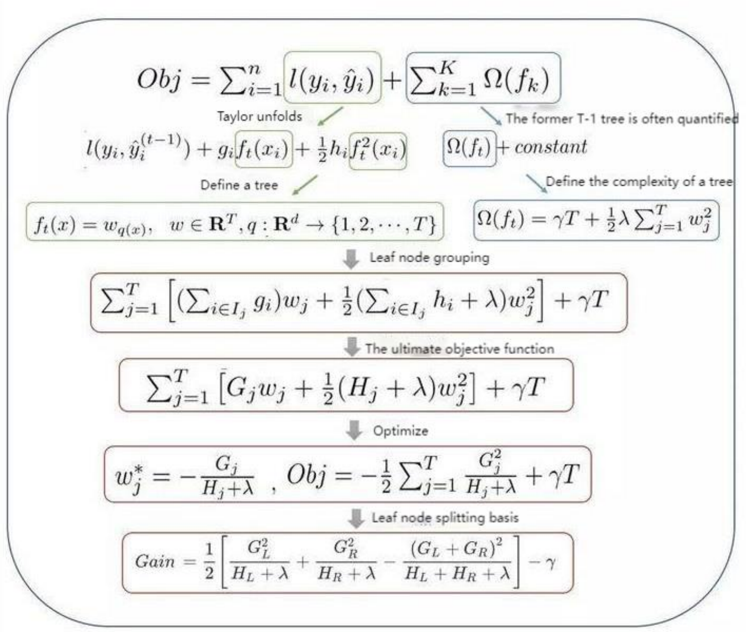


Figure 1. Schematic diagram of XGBoost derivation process

2.3. XGBoost discriminant prediction model flow chart

The XGboost flow chart is shown in Figure 2

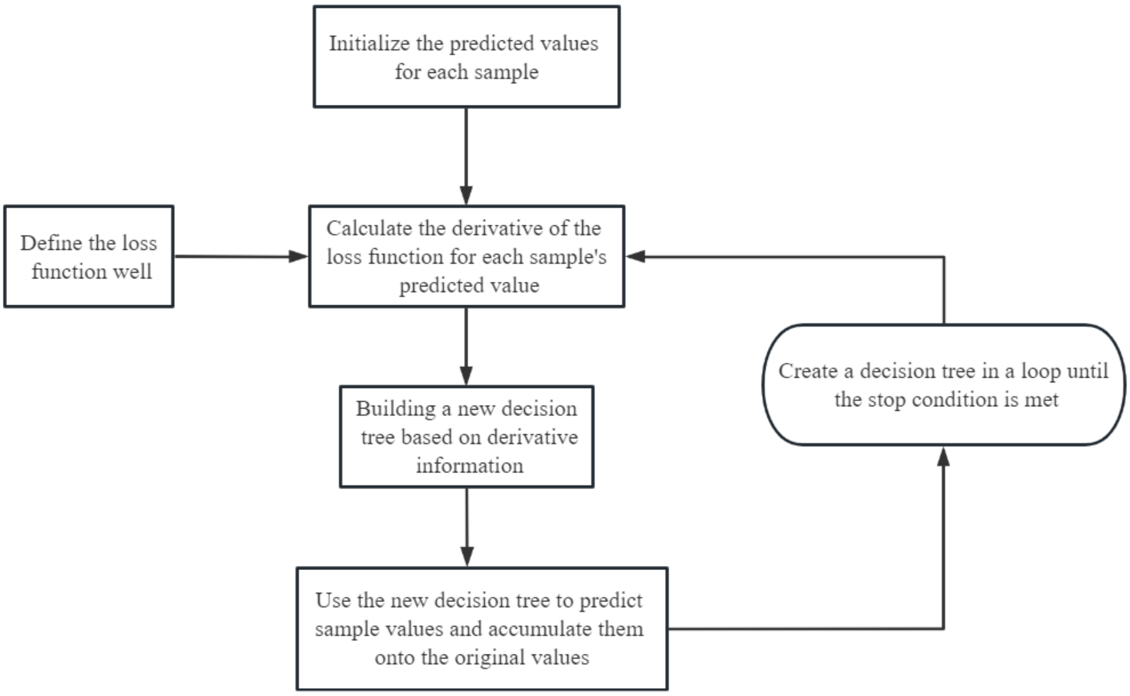


Figure 2. XGBoost flow chart

3. CRITIC analysis model

It is necessary to explore the main causes of each type of failure, find out the attributes related to them, and explore some rules contained in each type of failure. Considering that the factors affecting each type of failure are not necessarily the same, it is necessary to determine the main factors affecting each type of failure and analyze their internal patterns. Since it is necessary to determine the weight of each influencing factor, in order to objectively and effectively reflect the information covered by the data itself, this question uses the objective weighting method.

Among the objective weighting methods, the entropy weight method and the standard deviation method are the more frequently used methods in comprehensive evaluation methods. Compared with the entropy weight method and the standard deviation method, the CRITIC analysis method is a relatively better weighting method. It comprehensively measures the objective weight of indicators based on the contrast strength of evaluation indicators and the conflict between indicators [9], its advantage lies in taking into account the variability of the indicator while taking into account the correlation of the indicator.

3.1. Basic steps of CRITIC analysis model

Step 1: Based on data preprocessing, screen out the sample data of each fault according to the "specific fault category", and summarize and collect the selected sample data.

Step 2: Conduct CRITIC analysis on the sample data of each fault type in turn to obtain the weight of each influencing factor on each fault category.

Step 3: Determine the main influencing factors of each fault category and analyze the laws behind them based on production and life.

3.2. Brief description of CRITIC analysis method

CRITIC analysis is a method that fully utilizes the objective attributes of the data itself for scientific evaluation.

Step 1: Dimensionless processing of indicator data

In the evaluation process, indicators with larger indicator values and smaller effects are called negative indicators, and the rest are positive indicators. Therefore, it is necessary to perform positive processing and normalization processing on negative indicators, that is:

$$x_{ij}^{'} = \frac{x_{\max} - x_{ij}}{x_{\max} - x_{\min}} \quad (1)$$

Normalize the positive indicators, and the formula is:

$$x_{ij}^{'} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

Among them, x_{ij} is the normalized value of the j-th data of the i-th indicator; $x_{\min}$ is the minimum value in the i-th indicator data; $x_{\max}$ is the maximum value in the i-th indicator data.

Step 2: Calculate the variability of each influencing factor.

For the variability of each influencing factor, the CRITIC analysis method uses standard deviation to measure, that is, variability represents the difference and fluctuation of the internal values of each indicator. The larger the standard deviation, the greater the numerical difference of the indicator.

Step 3: Calculate the conflict of each influencing factor.

For the variability of each influencing factor, the CRITIC analysis method uses the correlation coefficient to measure it. The larger the correlation coefficient, the smaller the conflict between the influencing factor and other influencing factors, and the more the same information is reflected.

$$R_i = \sum_{i=1}^{m} (1 - r_{ij}) \quad (3)$$

Among them, r_{ij} represents the correlation coefficient between evaluation factors i and j.

Step 4: Calculate the information amount of each influencing factor.

$$d_i = S_i \sum_{i=1}^{m} (1 - r_{ij}) \quad (4)$$

Among them, d_i is the information amount of the i-th indicator; S_i is the standard deviation of the i-th indicator.

Step 5: Normalize to find the objective weight of each influencing factor.

$$w_i = \frac{d_i}{\sum_{i=1}^{m} d_i} \tag{5}$$

Among them, w_i is the entropy weight of the i-th indicator; d_i is the information amount of the i-th indicator.

4. Data verification

4.1. Missing value test

Use Python software to perform missing value tests on 10 indicators of the attached data and find that there are no missing values in the given data set.

4.2. Data analysis

According to the experimental data in the attachment, there are only 1-2 samples with the same number in the "Machine Number" feature, and there are no identical codes in the "Unified Standard Code" feature, which is of no significance for comparative analysis. Therefore, these two-feature data are deleted. deal with.

The "whether a failure occurs" and "specific failure category" indicators reflect the failure conditions of different machines. Therefore, when establishing a "whether a failure occurs" prediction model, it is necessary to use "whether a failure occurs" as the dependent variable and analyze the impact of other factors on it. ; When establishing a "specific fault category" discrimination prediction model, it is necessary to filter out the sample data of failure according to the "whether a fault occurs" feature data set, and then use the "specific fault category" as the dependent variable to explore the impact of other factors.

4.3. Outlier test

Use T test to detect outliers and find that there are no outliers. The specific steps are.

Step1: Calculate the mean and variance of each feature data, that is.

$$\begin{cases} \bar{x} = \sum_{i=1}^{n} x_i \\ \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \end{cases} \tag{6}$$

Among them, $\bar{x}$ is the mean value of each indicator column; n is the number of samples.

Step2: Use the largest error in the indicator data set as the tested quantity, if.

$$|x_m - \bar{x}| \geq t_a(n-1)\sqrt{1+\frac{1}{n-1}} \tag{7}$$

when, then discard x_m.

5. Fault inspection results

5.1. Solution of XGBoost discriminant prediction model

Step 1: Use the attached data as a training set, take "machine quality level", "room temperature", "machine temperature", "speed", "torque" and "usage time" as the input set, and take "whether a fault occurs" as the output set, establish the XGBoost discriminant prediction model.

Step 2: Calculate feature importance.

Step 3: Apply the established XGBoost model to the test set to evaluate the effect of the XGBoost discriminant prediction model.

Its solution parameters are shown in Table 1:

Table 1. XGBoost model parameter table

Parameter Name	Parameter value
Training time	6.822s
Data segmentation	0.7
data shuffle	Yes
cross validation	10
base learner	gbtree
Number of base learners	100
Learning rate	0.1
L1 Regular Term	0
L2 Regular Term	1
Sample sampling rate	1
Tree feature sampling rate	1
Node feature sampling rate	1
Minimum weight of samples in leaf nodes	0
The maximum depth of the tree	10

The importance of its features is shown in Figure 3:

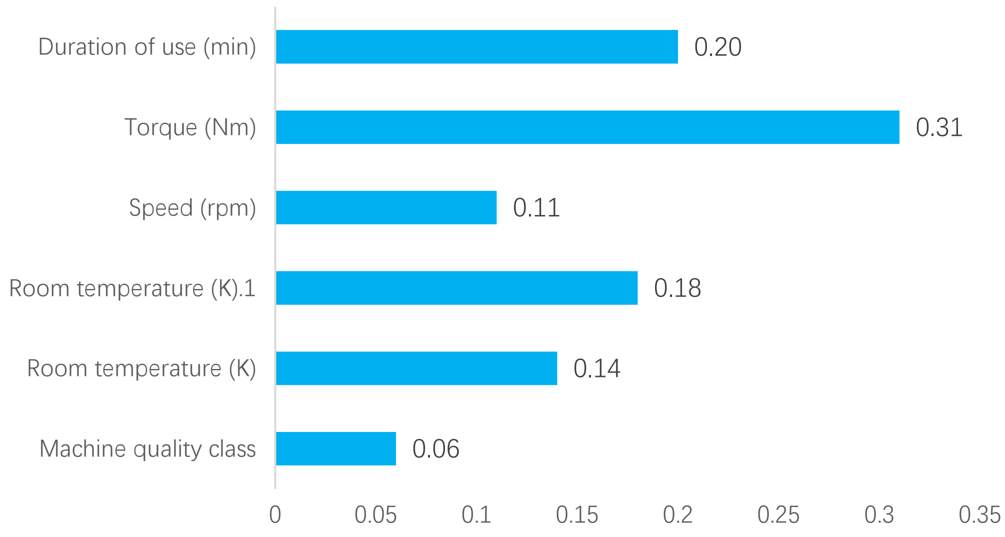
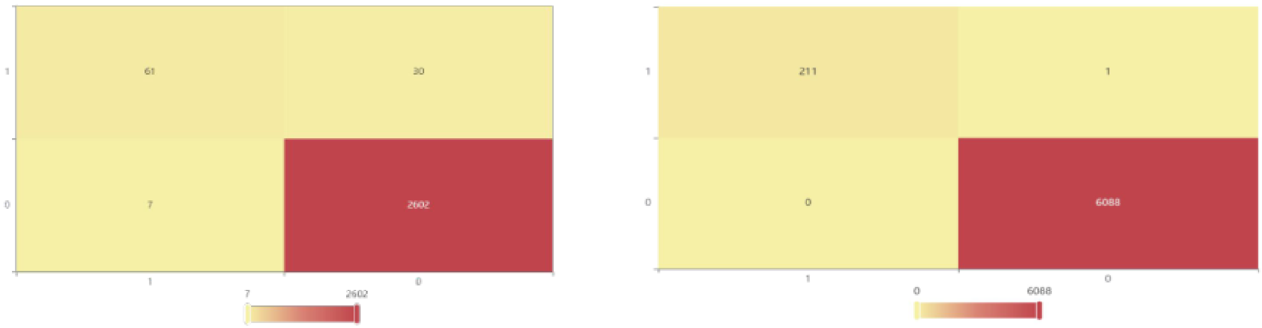


Figure. 3 Feature importance

Its confusion matrix heat map is shown in Figure 4:



(a)Test data heatmap (b)Training data heatmap

Figure. 4 Confusion matrix heat map

Since the machine learning algorithm is used to classify "whether a fault has occurred", the accuracy, recall rate, precision rate, and F1 index are used to evaluate the XGBoost discriminant prediction model. The results are shown in Table 2.

Table. 2 Model evaluation table

	Accuracy	Recall	Precision	F1
Training set	1	1	1	1
Cross validation set	0.984	0.984	0.983	0.983
Test set	0.986	0.986	0.986	0.985

The above table measures the prediction effect of XGBoost through quantitative indicators. Among them, the hyperparameters can be continuously adjusted through the evaluation indicators of the cross-validation set to obtain a reliable and stable model. Some of the test data tables are shown in Table 3:

Table. 3 Test data evaluation table (part)

forecast result	Has it occurred fault	Probability of predicted results_ 0	Probability of predicted results_ 1	Machine quality level	room temperature	Machine temperature	speed	torque	Usage duration
0	0	0.999846	0.000154	2	300.8	310.8	1461	41.3	53
0	0	0.999848	0.000152	2	299.9	309.8	1625	41.2	109
0	0	0.997387	0.002613	1	302.3	310.9	1403	47.2	173
0	0	0.999054	0.000946	2	303.3	311.8	1441	44.9	61
0	0	0.999704	0.000296	1	302.9	312.3	1515	40.8	183
0	0	0.999253	0.000747	1	299.6	310.8	1393	49.9	78
0	0	0.656857	0.343143	1	300.6	310.4	2514	13.4	215
0	0	0.971804	0.028196	1	298.9	309.7	1407	41.9	207
0	0	0.999603	0.000397	0	297.4	308.9	1645	30.8	196
0	0	0.999804	0.000196	0	299.8	309.2	1435	42.6	111
0	0	0.999597	0.000403	1	302	310.6	1467	44.9	86
0	0	0.998703	0.001297	1	299.4	308.8	1390	57	92
0	0	0.999121	0.000879	2	299.3	308.5	1304	60.6	8
0	0	0.999846	0.000154	1	297.9	309.1	1543	36.6	20
0	0	0.999852	0.000148	1	301.1	310.5	1532	37.5	119

5.2. Mechanical equipment failure prediction

The sample data in the experimental data were predicted based on the XGBoost discriminant prediction model, and it was found that a total of 19 mechanical equipment had faults. Based on the logistic regression discriminant prediction model, the 19 faulty devices were predicted and classified, and it was found that 4 of them were heat dissipation faults, 6 were power faults, and 9 were overload faults. The specific results are shown in Table 4:

Table. 4 Summary of faulty equipment

Fault category	number
TWF wear fault	0
HDF Heat dissipation failure	4
PWF Power failure	6
OSF Overload fault	9
RNF Other faults	0

The detailed information of the summary table is shown in Table 5:

Table. 5 Specific information table of faulty equipment

Machine no	Unified specification code	Is there a malfunction	Specific fault categories
9868	L47340	1	OSF
9553	L47341	1	OSF
9968	M15054	1	PWF
9587	L47422	1	OSF
9404	L47428	1	OSF
9618	L47429	1	OSF
9198	M15119	1	PWF
9887	L47507	1	OSF
9592	L47560	1	PWF
9922	L47643	1	PWF
9533	L47783	1	PWF
9970	L47926	1	OSF
9298	L48106	1	OSF
9507	M18674	1	HDF
9244	H33243	1	HDF
9057	L51260	1	HDF
9552	L51261	1	HDF
9737	L47230	1	PWF
9816	L47249	1	OSF

5.3. Analysis of causes of mechanical equipment failures

5.3.1 Weight analysis

According to the established CRITIC weight analysis model and the model steps, the weight calculation results of various factors affecting various types of faults are shown in Table 6:

Table. 6 Weight table of factors affecting failure.

	speed	torque	Usage duration	room temperature	Machine temperature	Machine quality level
TWF	0.900	0.048	0.041	0.005	0.004	0.002
HDF	0.326	0.062	0.597	0.005	0.005	0.005
PWF	0.881	0.041	0.073	0.002	0.002	0.001
OSF	0.722	0.076	0.170	0.017	0.012	0.003
RNF	0.717	0.037	0.237	0.004	0.003	0.002

5.3.2 Factors affecting wear failure.

The weight calculation results based on the CRITIC method show that the weight of machine quality grade characteristics is 0.241%, the weight of room temperature characteristics is 0.547%, the weight of machine temperature characteristics is 0.384%, the weight of speed characteristics is 89.966%, and the weight of torque is 4.757%, the weight of the usage time feature is 4.105%, where the maximum value of the index weight is the rotation speed feature, and the minimum value is the machine quality level feature.

5.3.3 Factors affecting heat dissipation failure.

The weight calculation results based on the CRITIC method show that the weight of the machine quality grade characteristics is 0.523%, the weight of the room temperature characteristics is 0.483%, the weight of the machine temperature characteristics is 0.527%, the weight of the rotational speed characteristics is 32.646%, and the weight of the torque characteristics is 6.169 %. The weight of the usage time feature is 59.651%. The maximum value of the indicator weight is the usage time feature, and the minimum value is the room temperature feature.

5.3.4 Factors affecting power failure.

The weight calculation results based on the CRITIC method show that the weight of machine quality grade characteristics is 0.068%, the weight of room temperature characteristics is 0.215%, the weight of machine temperature characteristics is 0.168%, the weight of speed (rpm) is 88.131%, and the weight of torque characteristics is 4.086%, and the weight of the usage time feature is 7.333%. The maximum value of the index weight is the rotation speed feature, and the minimum value is the machine quality level feature.

5.3.5 Factors affecting overload faults.

The weight calculation results based on the CRITIC method show that the weight of the machine quality grade characteristics is 0.321%, the weight of the room temperature characteristics is 1.691%, the weight of the machine temperature characteristics is 1.23%, the weight of the speed characteristics is 72.246%, and the weight of the torque characteristics is 7.555 %, the weight of the usage time feature is 16.957%, where the maximum value of the index weight is the speed feature, and the minimum value is the machine quality level feature.

5.3.6 Factors affecting other faults.

The weight calculation results based on the CRITIC method show that the weight of machine quality grade characteristics is 0.158%, the weight of room temperature characteristics is 0.426%, the weight of machine temperature characteristics is 0.325%, the weight of speed is 71.714%, and the weight of torque characteristics is 3.719%, the weight of the usage time feature is 23.657%, where the maximum value of the index weight is the rotation speed feature, and the minimum value is the machine quality level feature.

6. Conclusion

Through in-depth analysis of mechanical equipment failure data and model establishment, the following conclusions are drawn rotation speed characteristics play a vital role in mechanical equipment failures, not only the main cause of wear failures, but also related to power failures, overload failures, etc. Various fault types are closely related. In addition, age characteristics also play a key role in thermal failures.

In actual industrial production, it is recognized that wear failures are closely related to factors such as component friction, vibration, and fatigue. Heat dissipation failures are mainly related to factors such as heat accumulation caused by high-speed operation, increased friction resistance, rising ambient temperature, and poor ventilation. The occurrence of power failure is limited by the number of equipment that the company can operate at the same time, while overload failure is usually caused by exceeding the maximum load that the mechanical equipment can withstand.

Therefore, it is recommended that during the use of mechanical equipment, special attention be paid to the rotation speed and working status of the mechanical equipment, and the temperature of the equipment is monitored at the same time. This helps to identify potential signs of failure early and take preventive measures or emergency treatment, thus improving the reliability and economic efficiency of production equipment. This research provides enterprises with practical fault prediction and prevention strategies and is expected to be widely used in the manufacturing industry to improve the stability and efficiency of equipment operation.

References

[1] Fan Guodong, Li Bohan. Industrial machinery equipment fault prediction method based on machine learning [J]. Automation and Information Engineering, 2023, 44(04): 13-18+50.

[2] Tian Hui, Liu Weibin. Research on diesel engine failure prediction based on single classification algorithm SVDD [J]. China Water Transport (Second Half), 2022, 22(09): 70-72.

[3] Huang Chen. Research on shield equipment fault fusion prediction model based on ensemble learning [D]. Southwest Jiaotong University, 2022.

[4] Liang Chen. Application of artificial neural networks in industrial equipment failure prediction[D]. Shanghai University of Finance and Economics, 2022.

[5] Zhang Jian, Wang Wenquan, Liu Wenbiao, etc. Mechanical equipment fault data enhancement method based on generative adversarial network [J]. Industrial Control Computer, 2021, 34(12): 108-109+111.

[6] Wang Chunzhu. Research on equipment failure prediction method based on residual LSTM and width learning [D]. Beijing University of Posts and Telecommunications, 2022.

[7] Huang Pengcheng. Research on hydraulic system fault prediction algorithm based on multi-task deep learning [D]. Guangdong University of Technology, 2022.

[8] Lin Like, it's so late to meet you! One article to understand the XGBoost algorithm, https://blog.csdn.net/qq_40589204/article/details/115481924. October 30, 2022.

[9] Yang Nan, Jin Aibing, Wang Jie, Chen Shuaijun, Jiang Linjing. Optimization of stope parameters at different depths based on CRITIC-TOPSIS comprehensive evaluation [J]. Mining Research and Development, 2022, 42(10): 6-12