1. Nataliia LUTSKA[1], 2. Nataliia ZAIETS[2], 3. Lidiia VLASENKO[3]

National University of Food Technologies (1), National University of Life and Environmental Sciences of Ukraine (2), State University of Trade and Economics (*3)*

ORCID: 1. 0000-0001-8593-0431; 2. 0000-0001-5219-2081, 3. 0000-0002-2003-6313

doi:10.15199/48.2024.02.33

Development of Diagnostic System for the State of Electric Drives of Food Enterprise

***Abstract**. The paper presents a system for diagnosing the state of electric drives based on a machine learning classification model. At the initial stage, the behavior of the diagnostic system and external subsystems was modelled based on UML diagrams. The input data for the model for diagnosing the state of electric drives is a set of data, which, as a rule, are measured in food enterprises. The deterioration of qualitative classification scores with an incomplete feature vector was studied when using different training methods.*

***Streszczenie.** W artykule przedstawiono system diagnozowania stanu napędów elektrycznych oparty na modelu klasyfikacyjnym uczenia maszynowego. W początkowej fazie zamodelowano zachowanie systemu diagnostycznego oraz podsystemów zewnętrznych w oparciu o diagramy UML. Danymi wejściowymi do modelu diagnozowania stanu napędów elektrycznych jest zbiór danych, które z reguły są mierzone w przedsiębiorstwach spożywczych. Zbadano pogorszenie jakościowych wyników klasyfikacji z niepełnym wektorem cech przy użyciu różnych metod szkoleniowych. (**Opracowanie Systemu Diagnostyki Stanu Napędów Elektrycznych Przedsiębiorstwa Spożywczego**)*

Keywords: *electric drives*, *system for diagnosing, classification, quantitative estimates, scenarios*
Słowa kluczowe: *napędy elektryczne, system diagnozowania, klasyfikacja, szacunki ilościowe, scenariusze*

Introduction

The transition of food production to asset-oriented management subsystems in accordance with the concept of Industry 4.0 [1, 2] requires the integration and redistribution of all types of production support. At the same time, the automated production control system becomes decisive, which becomes an integral part of any subsystem. First, it includes an automated process control system, which is the primary source of information about the state of the process and equipment. In addition, the automated production control system ensures the coordination and implementation of the feedback of these subsystems, thus realizing the functions of MES/MOM [3, 4]: operational management of product development and maintenance, management of product quality assurance operations, inventory management. Maintenance operations management systems require developers to diagnose the state of equipment and technical devices [5], plan operational, periodic, preventive or advanced maintenance. Such systems should provide a systematic approach to the problem, taking into account both the typical nature of the equipment and the peculiarities of its functioning in the corresponding plant.

Every year, the requirements for the reliability of technical means [6] increase, and the efforts applied to improve them are insufficient to meet the new challenges of the time. The following primary reasons can be identified that determine the need to increase attention to the problems of reliability of technical means of food production:

- increasing the complexity of systems that include a large number of individual components and elements;
- increasing the importance of the functions performed by the components of the systems;
- complication of operating conditions of systems;
- lagging of the growth rate of the reliability of components from the growth rate of the number of components in the systems;
- the intensity of the operating modes of the system or individual nodes;
- increasing requirements for the quality of equipment operation;
- an increase in the responsibility of the functions performed by the system, the high economic and technical cost of failure;
- full or partial automation and, as a result, the exclusion of direct human control of the functioning of the system and its elements.

The main requirement for the operation of any technological complex of food production is to obtain the highest profit, which is the result of the continuous operation of all interacting systems – technological, electrical and automated control systems. This is possible only in the absence of downtime, compliance with the technological regulations, timely detection of deviations of the controlled variables from the set values, as well as the absence of breakdowns of technological equipment and technical means. The latter are determined by the corresponding technical condition, characterized by the requirements established by the regulatory and technical documentation. Therefore, increasing the reliability of automation technical means is an urgent task, and the development of a system for diagnosing the state of technical devices and predicting their failure is becoming an indispensable part of modern production.

Special attention is paid to electrical equipment, since their uninterrupted operation ensures the coordination of adjacent production departments. Therefore, it is important to determine their condition and identify the need for repair work at an early stage. Engine breakdowns can be detected by various methods, for example, based on data analysis from fault recorders [7], vibration signal analysis and artificial neural networks [8], expert assessments [9], comparison of simulation results with calculated device operation characteristics, different experimental methods [10].

Both supervised and semi-supervised techniques and unsupervised techniques can be used to detect breakdowns [11-14]. Each of them has its advantages and disadvantages; however, preference is given to one or another method, taking into account the characteristics of the data and the requirements for the task. The work uses a variety of supervised learning methods that have proven to be reliable, less sensitive to outliers. Such a drawback as the reflection of the full range of breakdowns in the data set is eliminated during operation by collecting new data for training and retraining the model.

For recognition systems, machine learning methods have proven themselves well [15, 16], using different metrics: probabilities (Naive Bayes Classifier, Linear or

Quadratic Discriminant Analysis); distances (Nearest Neighbor Method, k-Nearest Neighbors, Support Vector Machines); projections (Linear or Logistic Regression); uncertainties (Decision Tree, Random Forest, etc.). However, the choice of the final method is again the problem of the designer, guided by the characteristics of the data collected itself.

Methods

In the structure of the electrical complex of food production in Ukraine, asynchronous AC motors with different powers and speeds are mainly used. According to the functions performed, electric motors in the food industry can be divided into electric drives for pumps; electric drives for shaft rotation (mixers, screws, etc.); damper actuators.

From their reliability, first, depends the successful operation of all technological equipment of the enterprise, and as a result, the productivity and quality of the finished product. Despite this, food industry enterprises pay more attention to the components of technological equipment and do not control the state of electrical machines between scheduled repairs. Electric drive problems often do not reveal themselves until the actual failure or are simply difficult to detect without the use of specialized diagnostics. Untimely maintenance or simply the lack of any control over the technical condition of electrical machines can lead not only to their failure, but also to accidents in production.

To develop a system for diagnosing and predicting breakdowns of such tools, a diagram of the system behavior was created (Fig. 1). In particular, the diagram in Fig. 1a visualizes the behavior of all subsystems of the system under consideration, taking into account the fact that the diagnostic system interacts with the archive of technological variables. From the archive, it receives data on the values of the relevant process variables (obtained from the automation system, technology laboratory or calculated using the appropriate formulas). The diagnostic system, based on the embedded algorithms [17, 18], determines the presence of a breakdown of electrical equipment and, on this basis, forms an application for the repair team to carry out work. After receiving the application, the repair team inspects the breakdown, determines its complexity and the amount of work to eliminate it. This can be repair for a complex breakdown and maintenance for a minor problem. At the end, the team reports on the performance of the action. When the operability of a unit of electrical equipment is restored, the order is considered completed and is closed.

Modeling the behavior of the element (Fig. 1b) reflects how the diagnostic system interacts with the repair team: it places a request for repair work. The system itself monitors the condition of electrical equipment and the occurrence of breakdowns. The behavior of the system is modeled by several use cases: general and variable. For each of these use cases, a behavior specification is provided.

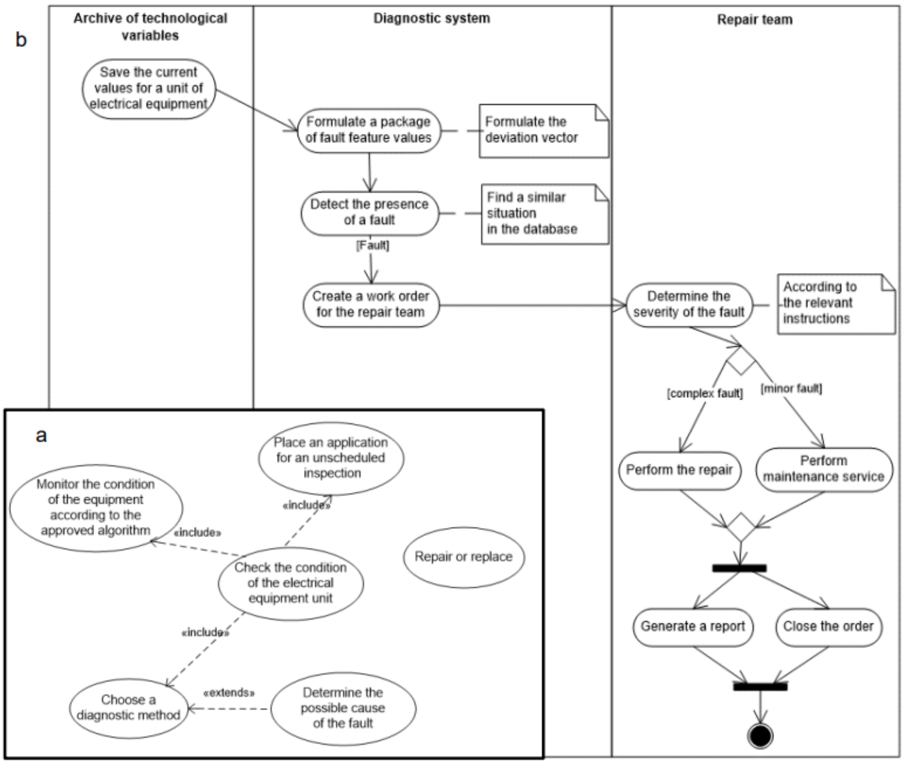

Fig.1. UML-diagrams of the behavior of the system for diagnosing the state of electric drives: a – Use Case; b – Activity

The fleet of electrical equipment at food enterprises in Ukraine has more than a hundred asynchronous motors of various capacities. To create an intelligent system for predicting the reliability of electric motors, it is necessary to divide them into classes not only according to the functions performed, but also depending on the power and speed.

Thus, the mathematical model for predicting breakdowns of the electric drive can be described by the following relationship:

$$\hat{h}(\mathbf{x}) = \operatorname{argmax}_c P(t = c \mid \mathbf{x}, \boldsymbol{\theta}) \quad (1)$$

where: $\hat{h}(\mathbf{x})$ – engine breakdown forecast; P – the probability of breakage; t, c – target variable (breakdown) and the class to which it was assigned $\{0,1\}$; $\boldsymbol{\theta}$ – the vector of model parameters; $\mathbf{x}$ – the space of measuring variables.

At different productions of the food industry, different variables (vector $\mathbf{x}$) can be measured that affect the final forecast. The authors have allocated a space of features that determine the current state of the electric drive:

1 – Average load, % – x_1;
2 – Rotational frequency, rpm – x_2;
3 – Power supply voltage, V – x_3;

4 – Current consumption, A – x_4;
5 – Leakage current, mA – x_5;

The first stage of the study is to build a feature-based machine learning model using data collected from a typical medium-capacity sugar mill for electric drives with power of 55 kW, 1500 rpm. A total of 2800 samples were collected.

The scatterplot shown in Fig. 2 shows a good separation of the two classes, so we use the following training methods: logistic regression; quadratic discriminant analysis; naive bayes classifiers; support vector machine; binary decision tree; classification neural network. As noted, they are based on different metrics: the first one is based on projection metrics; the second and third are based on probabilities; the fourth is based on distances; the fifth is based on uncertainty; the sixth is based on any of the named metrics.

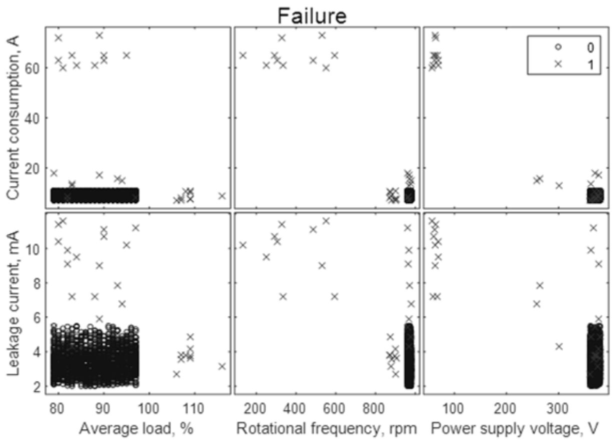

Fig.2. Matrix of scatter plots

In addition, a feature of such a data set is their skewed classes; in particular, the amount of data on the normal operation of engines is > 95% of the entire sample. This requires careful selection of model evaluation criteria. To compare the effectiveness of the selected methods, the data set is divided into two samples - training and test, for each of which the following indicators are calculated:

- F_β-score with β = 0.1 (F_β);
- Matthews correlation coefficient (MCC);
- Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

F_β-score balances precision and recall with the benefit of recall. MCC provides for various types of errors and an objective assessment in the face of imbalanced data. AUC-ROC evaluates the overall ability of a model to distinguish between classes without being tied to a specific threshold.

At the second stage of the study, we consider the possibility of building a model (1) using a different combination of input variables from the vector $\mathbf{x}$. Thus, it will be possible to generalize the results, that is, to determine the minimum number of input variables and the classification method for an acceptable level of recognition of the class of electric drive failure.

Results

A lot of machine learning models were built for all five features $\mathbf{x}$ – the resulting estimates with the best architectures and hyperparameter optimization are shown in Table. 1. All the obtained models have high performance both on the training and test sets and can be used to predict the breakdowns of electric drives.

Consider options for obtaining a forecast model in the absence of some features from the vector $\mathbf{x}$. Of the reasonably possible scenarios, 17 were selected with different combinations of features (Fig. 3). At least two of the five features need to be measured. Leakage current x_5 was not included only in two – 26 and 30, and in the 26th there are all other signs [x_1, x_2, x_3, x_4], and in the 30th - three [x_1, x_2, x_4]. Thus, x_5 is the main sign of failure of the engines under study. However, if it is impossible to register it, it is possible to replace it with others.

Table 1. Quantitative estimates of the constructed models

Method	F_β Train	F_β Test	AUC Train	AUC Test	MCC Train	MCC Test
Logistic regression	0.971	0.882	1.000	0.985	0.982	0.925
Quadratic Discriminant Analysis	0.971	1.000	1.000	1.000	0.982	1.000
Naive Bayes Classifiers	0.971	1.000	1.000	1.000	0.982	1.000
Support vector machine	0.971	0.882	1.000	0.997	0.982	0.925
Decision Tree	1.000	1.000	1.000	1.000	1.000	1.000
Neural Network	0.971	1.000	1.000	1.000	0.982	0.925

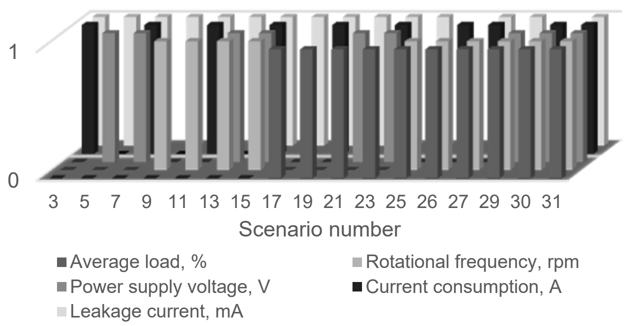

Fig.3. Presence of features by scenarios

Consider the obtained estimates of the quality of models for the selected scenarios Fig.4-6. We take into account that the value >0.7 of all three estimates will provide an acceptable quality of model classification, and with a value of >0.9, more stringent requirements are provided for the accuracy and ability of the model to classify data. The binary decision tree method showed the highest performance in all scenarios. We have the best forecasting quality (all estimates are maximum) for the 9th, 11th, 13th, 15th, 17th, 19th, 21st, 23rd and 25th scenarios when using the binary decision tree method, moreover, on 3rd, 5th, 7th, 26th and 30th – all scores >0.75. The specified set of scenarios assumes the presence of the feature x_5. Thus, a method that uses an uncertainty metric for such a distribution is the best.

The Logistic regression method on scenarios 9, 11, 13, 15, 25, 26, 27, 29, 30 and 31 has >0.7 scores, and >0.9 scores on scenarios 13 and 29. That is, when using this method, the presence of the feature x_2 is necessary. The methods of Quadratic Discriminant Analysis and Naive Bayes Classifiers on scenarios 9, 11, 13, 15, 25, 26, 27, 29, 30 and 31 have scores >0.7, and on 11, 15, 27 and 31 – >0.9. Quadratic Discriminant Analysis also scores >0.7 on scenarios 13 and 17. The SVM method has scores >0.9 for the 11th, 15th, 25th, 26th, 27th, 30th, and 31st scenarios, which include at least three features, of which the presence of x_2 is mandatory. The Neural Network method has the worst scores (<0.7) on scenarios 3, 5, and 11, while the best scores (scores >0.9) are on 9th, 19th, 25th, and 31st, that is, in the presence of x_2 and x_5 features.

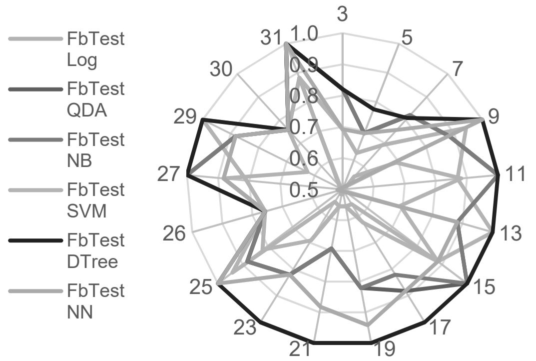

Fig.4. Estimation of F_{β}-score of different models by scenarios

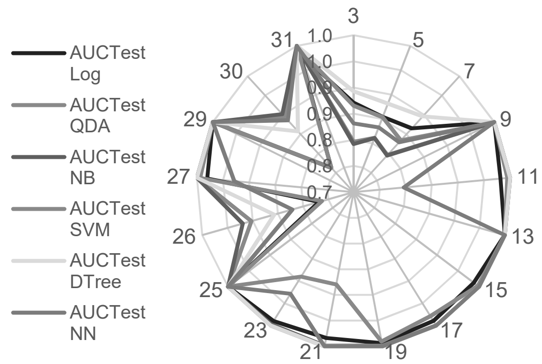

Fig.5. AUC-ROC Estimation of Different Models by Scenarios

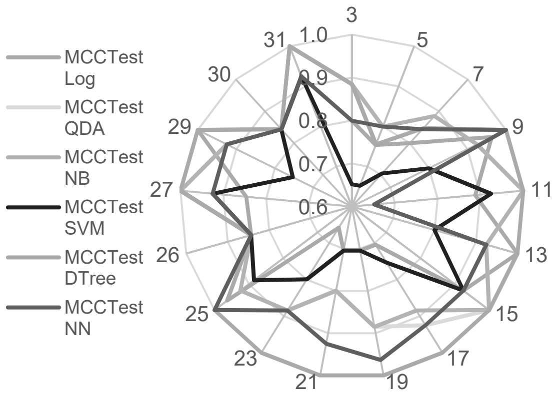

Fig.6. Estimation of MSS of different models by scenarios

If we evaluate the general trend by methods, then scenario 11, which provides for the feature vector [x_2, x_4, x_5], when using the Quadratic Discriminant Analysis, Naive Bayes Classifiers, Support vector machine and Decision Tree methods, has all scores > 0.88. In scenario 15 with features [x_2, x_3, x_4, x_5], the Neural Network method was added to the above methods. For the 3rd and 5th scenarios, when the features [x_4, x_5] and [x_3, x_5] are used, respectively, all estimates are >0.7 only when using the Decision Tree method. It can also be concluded that the SVM method only for 7 out of 17 scenarios has all scores >0.7.

Conclusions

The paper proposes a system for diagnosing and predicting electric drives based on machine learning methods. Evaluation data of electric drives on normal operation and breakdowns was collected from a sugar factory in Ukraine for several seasons of its operation. They were trained on different types of supervised models based on different metrics. Variants of scenarios have been developed that provide for the presence of an incomplete feature vector and recommendations have been issued on the use of the training method. The highest performance in all scenarios was achieved using the binary decision tree method.

***Authors**: DscTech, Prof. Nataliia Lutska, National University of Food Technologies, Volodimirska str., 68, 01601, Kyiv, Ukraine, E-mail: lutskanm2017@gmail.com; DscTech, Prof. Nataliia Zaiets, National University of Life and Environmental Sciences of Ukraine, Defense Heroes str., 15, 03041, Kyiv, Ukraine, E-mail: z-n@ukr.net; PhD, Assoc Prof. Lidia Vlasenko, State University of Trade and Economics, Kyoto str., 19, 02156, Kyiv, Ukraine, E-mail: vlasenko.lidia1@gmail.com.*

REFERENCES

[1] Wagner C., et al. The role of the Industry 4.0 asset administration shell and the digital twin during the life cycle of a plant, *22nd IEEE international conference on emerging technologies and factory automation (ETFA)*. IEEE, (2017), 1-8
[2] Nam V.Q., Huy D.T.N., Hang N.T., Le T.H., Thanh N.T.P., Internet of Things Effects and Building Effective Management Information System, *Webology*, (2021), 354-363
[3] Mantravadi S., Møller C., Chen L.I., Schnyder R., Design choices for next-generation IIoT-connected MES/MOM: An empirical study on smart factories, *Robotics and Computer-Integrated Manufacturing*, 73 (2022), 102225
[4] Mantravadi S., Møller C., An Overview of Next-generation Manufacturing Execution Systems. *Procedia Manufacturing*, 30 (2019), 588-595
[5] Straka L., Corny I., Krehel R., Evaluation of capability of measuring device on the basis of diagnostics, *Applied Mechanics and Materials*, 308 (2013), 69-74
[6] Duer S., Rokosz K., Zajkowski K., Bernatowicz D., Ostrowski A., Woźnia, M., Iqbal A., Intelligent Systems Supporting the Use of Energy Devices and Other Complex Technical Objects, *Energies*, 15 (2022), No. 17, 1-6
[7] Sajewicz D., Łaguna W., Chmielak W. Detection of fault events in electric motor based on analysis of data from fault recorders, *Przegląd Elektrotechniczny*, *99* (2023), No. 1, 207-210
[8] Ribeiro Junior R.F., de Almeida F. A., Gomes G. F., Fault classification in three-phase motors based on vibration signal analysis and artificial neural networks, *Neural Computing and Applications*, *32* (2020), 15171-15189
[9] Zhongming Y.E., Bin W.U., A review on induction motor online fault diagnosis, *Proceedings IPEMC 2000. Third International Power Electronics and Motion Control Conference,* (2000), No. 3, 1353-1358
[10] Aimer A.F., Boudinar A.H., Khodja M.A., Bendiabdellah A., Frequency resolution improvements in induction motor fault diagnosis: Experimental validation, *Przeglad Elektrotechniczny*, 97 (2021), No. 11, 69-73
[11] Jiang T., Gradus J.L. Supervised machine learning: a brief primer, *Behavior Therapy*, 51 (2020), No. 5, 675-687
[12] Zheng M., You S., Huang L., Wang F., Qian C., Xu, C., Simmatch: Semi-supervised learning with similarity matching, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* (2022), 14471-14481
[13] Krishnan R., Rajpurkar P., Topol E.J., Self-supervised learning in medicine and healthcare, *Nature Biomedical Engineering*, 6 (2022), No. 12, 1346-1352
[14] Kang Z., Catal C., Tekinerdogan B., Machine learning applications in production lines: A systematic literature review, *Computers and Industrial Engineering*, 149 (2020), 106773
[15] Wuest T., Weimer D., Irgens C., Thoben K.D., Machine learning in manufacturing: Advances, challenges, and applications, *Production and Manufacturing Research*, 4 (2016), 23-45
[16] Bulusu S., Kailkhura B., Li B., Varshney P.K., Song D., Anomalous Instance Detection in Deep Learning: A Survey, arXiv:2003.06979, (2020)
[17] Vlasenko L., Lutska N., Zaiets N., Korobiichuk I., Hrybkov S., Core Ontology for Describing Production Equipment According to Intelligent Production, *Appl. Syst. Innov.,* 5 (2022), No. 5, 98
[18] Lutska N., Vlasenko L., Zaiets N., Lysenko V., Modeling the Productivity of a Sugar Factory using Machine Learning Methods, *IEEE 17th International Conference on Computer Science and Information Technologies*, (2022), 34-43