ANNOTATION
to the dissertation work of Vladimir Anatolyevich Madin on the topic «Development and research of Big Data mining models in task planning» submitted for the degree of Doctor of Philosophy (PhD) in educational program 8D06102 – «Information technologies and robotics»

The relevance of this research topic is determined by current trends in the digital transformation of industry and the development of intelligent manufacturing systems based on big data analysis, machine learning, and optimization modeling. In the context of Industry 4.0, manufacturing systems increasingly function as cyber-physical systems, in which data from monitoring equipment, production processes, and technological operations becomes a key resource for improving management and decision-making efficiency.

The development of digital manufacturing platforms is accompanied by a shift from traditional scheduling methods to intelligent decision support systems capable of accounting for the uncertainty of production process parameters, variability in operation durations, and the technical condition of equipment. Production scheduling is viewed as part of an integrated production management system, including data collection, forecasting process parameters, and adapting production schedules.

In the Republic of Kazakhstan, the digitalization of industry, the implementation of intelligent information systems, and the development of data analysis technologies are considered priority areas of scientific and technological development. Under these conditions, the need for intelligent production process planning methods to improve the resilience of production systems and the efficiency of equipment utilization is growing.

A classical mathematical model for such problems is the Job Shop Scheduling Problem (JSSP). Despite significant progress in discrete optimization methods and scheduling theory, most existing scheduling models are based on deterministic parameter values and do not account for the uncertainty of production processes.

In real-world enterprise operations, operation durations, equipment utilization, and failure rates are subject to significant fluctuations, reducing schedule stability and requiring the development of planning methods capable of accounting for the uncertainty of production process parameters. At the same time, machine learning methods are rapidly developing to predict production system parameters and equipment performance based on monitoring data analysis.

Predictive maintenance (PdM), which focuses on forecasting the remaining useful life of equipment based on sensor data, is becoming increasingly important in modern production systems. Despite significant research in machine learning and predictive maintenance, predictive models are most often used in isolation from scheduling tasks, and their formal integration into optimization models remains underdeveloped.

Thus, three interconnected research areas have emerged in modern scientific and engineering practice: scheduling of production processes, machine learning methods for predicting production system parameters, and predictive equipment

maintenance methods. Integrating these areas within a unified computational architecture for intelligent planning represents a pressing scientific challenge.

The need to develop methods that ensure formalized integration of predictive machine learning models, predictive equipment maintenance data, and optimization scheduling procedures in a single computing circuit determines the scientific problem of this study.

The aim of the doctoral dissertation is to develop and validate a set of big data analytics models for predictive planning of production activities, aimed at increasing the validity and effectiveness of management decisions in conditions of uncertainty in the parameters of production processes.

Objectives of the dissertation research:

- analyze modern methods of data mining and machine learning applied to forecasting and scheduling tasks in production systems;
- develop a conceptual model for integrating heterogeneous production and operational data, as well as external factors considered as expandable sources of information in predictive planning tasks;
- construct hybrid forecasting models with a formalized procedure for assessing forecast uncertainty;
- develop an optimization model for the formation of a production plan considering scenario constraints on risks and quantile forecast estimates of operation parameters;
- integrate forecast models and predictive maintenance data into a single planning framework;
- conduct experimental testing of the method on open data sets with a comparison of the results with basic methods based on the metrics of forecasting accuracy, decision stability, and computational efficiency.

Object of the study: processes of scheduling production activities under conditions of uncertainty of the parameters of operations and the state of equipment.

Subject of the research: Hybrid predictive-optimization models for scheduling production activities, based on the integration of predictive machine learning models and optimization methods through the formation of scenario constraints that consider the uncertainty of production process parameters.

Research methods. At various stages, this dissertation the research employed methods of system analysis and mathematical modeling of complex production systems, machine learning and data mining, optimization modeling and combinatorial optimization, and mathematical statistics to assess forecast accuracy, the robustness of the resulting solutions, and the verification of computational experiment results.

Provisions submitted for defense:

- A hybrid predictive-optimization scheduling model based on the use of quantile forecasts of production process parameters and ensuring the integration of machine learning predictive model results into optimization scheduling models to account for the uncertainty of operation durations;
- A mathematical scheduling model considering the risks of the technical condition of equipment, based on the use of scenario-based constraints formed

following the results of the analysis of predictive maintenance indicators of equipment;

- An algorithm for the integration of predictive machine learning models and optimization planning, implementing the computational loop «forecasting – constraint synthesis – optimization – stability analysis» and ensuring an increase in the stability of the generated production schedules.

Key research findings. The study resulted in the development of a hybrid predictive-optimization approach to solving scheduling problems based on the integration of machine learning, predictive equipment maintenance, and discrete optimization methods. An intelligent scheduling system architecture is proposed that combines quantile predictive models of operation durations, equipment PdM risk assessment, and an optimization scheduling model based on the CP-SAT solver.

A machine learning model has been developed for generating quantile forecasts of operation durations (P10, P50, P90), allowing for the uncertainty of production processes to be considered when formulating optimization constraints. A method for integrating equipment technical condition indicators into a scheduling problem based on aggregated PdM risks is proposed.

An optimization model for scheduling with adaptive scenario constraints was developed, implementing planning scenarios S0-S5 and a mechanism for sequentially relaxing ML constraints in the absence of a feasible solution. A software prototype of an intelligent planning system was created, enabling forecasting of production process parameters, scheduling, and analytical visualization of results.

Computational experiments conducted on benchmark instances of the JSSP task confirmed the viability of the proposed approach and showed that the use of predictive machine learning models as part of the ML-CP-SAT pipeline increases the stability of the optimization process and ensures the formation of feasible schedules under uncertainty.

The scientific novelty lies in the development and validation of a hybrid predictive-optimization approach to production scheduling based on the integration of machine learning models and constraint-based optimization methods. A methodology for generating scenario constraints for an optimization model based on quantile forecasts of operation durations has been developed to account for the uncertainty of production process parameters. A method for integrating equipment technical condition assessments obtained using predictive maintenance methods into the scheduling framework has been developed. The proposed approach implements a unified computational framework of «forecasting - constraint synthesis - optimization - stability analysis», ensuring the reproducibility of computational experiments and the interpretability of planning decisions.

Theoretical significance. The aim of this study is to develop methods for integrating machine learning models and optimization procedures for scheduling production processes under uncertainty. A formalized framework for integrating predictive models and equipment condition assessments into an optimization formulation of the scheduling problem is proposed. This framework utilizes quantile estimates of operation parameters and expands the robust scheduling methodology for production processes.

Practical significance. The project involves developing a software prototype for an intelligent scheduling system that integrates machine learning, predictive equipment maintenance, and optimization modeling. The proposed methods for hybrid predictive-optimization scheduling can be used in the creation and modernization of decision support systems for production process scheduling, as well as in educational settings for training specialists in data analysis and intelligent information systems.

Implementation of research results. The results of this dissertation have been implemented in practice at OPTISORT LLC to improve the validity and quality of decisions in the area of production process scheduling through the use of predictive models and scenario analysis.

Validation of the dissertation research results. The main results of the dissertation research were presented and discussed at scientific seminars of the Software Department of the NJSC «Akhmet Baitursynuly Kostanay Regional University», as well as at the international webinar «Methods of Artificial Intelligence», Kostanay, 2022.

One article in international journals with a non-zero impact factor included in the Scopus database. The author's contribution includes formulating the scientific hypothesis, collecting and preprocessing the data, developing and implementing the experimental portion of the study, interpreting the results, and preparing the publication.

1. Enhancing electricity consumption forecasting in the Republic of Kazakhstan using machine learning // Journal of Applied Engineering and Technological Science. – 2025. – Vol. – 6(2). – P. 1166-1196. https://doi.org/10.37385/jaets.v6i2.7425.

Articles in journals recommended by the Committee for Quality Assurance in Science and Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan – 6 articles. In these publications, the author developed the research concept and data analysis methodology, participated in data collection, interpretation of results, and preparation of scientific articles for publication:

1. Convolutional and deep neural networks // Bulletin of KazATK. – 2023. – No. 126 (3) pp. 189-197. https://doi.org/10.52167/1609-1817-2023-126-3-189-197.

2. Comparative analysis of protocol formats for data processing in sensor networks of the Internet of Things // Bulletin of KazATK. – 2024. – No. 134 (5) pp. 465-473. https://doi.org/10.52167/1609-1817-2024-134-5-465-473.

3. Adaptive process management using deep learning on a programmable logic controller (PLC) // International journal of information and communication technologies. – 2024. – Vol. 5. Is. 1. No. 17 R. 8-28. https://doi.org/10.54309/IJICT.2024.17.1.001.

4. Analysis of RNN and TRANSFORMER models in natural language processing tasks // Vestnik KazATK. – 2025. – No. 138 (3) pp. 311-321. https://doi.org/10.52167/1609-1817-2025-138-3-311-321.

5. Predictive maintenance of spindle bearings based on vibration diagnostics (using the CKE 6136/6150 lathe as an example) // Vestnik KazUTB. – 2025. – Vol. 3. Issue 28. Pp. 63-79.

6. Dynamic reconfiguration of control algorithms in a programmable logic controller for flexible automation of production processes // VKTU Bulletin. – 2025. – No. 4 (110) pp. 174-185.

The author's personal contribution includes formulating the research problem, developing the concept of a hybrid predictive-optimization approach to scheduling, and analyzing the scientific literature on scheduling, machine learning, predictive equipment maintenance, and discrete optimization. The author developed the architecture of an intelligent scheduling system, quantile forecasting models for operation durations, a method for integrating PdM equipment risks into an optimization model, and a scenario-based mechanism for applying ML constraints. The author also developed the software implementation of the system prototype, prepared and processed data, conducted computational experiments, and analyzed the results. All the main results of this dissertation were obtained by the author personally.

Publications on the topic of the dissertation research. Seven (seven) scientific papers have been published on the topic of the dissertation research, including six (six) articles in journals recommended by the Committee for Quality Assurance in Science and Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan, and one (one) article in an international scientific journal included in the Scopus database. Two (two) copyright certificates for state registration of the computer program have also been granted.

Structure and scope of the dissertation. The dissertation consists of an introduction, five sections, a conclusion, a list of references, and appendices. The dissertation is 137 pages long (38 figures and 17 tables).

The introduction substantiates the relevance of the research topic, which is related to the need to improve the efficiency of production scheduling in the face of uncertain operation durations and equipment failure risks. The importance of applying machine learning, predictive maintenance, and discrete optimization methods to intelligent decision support in production planning problems is demonstrated. The purpose and objectives of the study are formulated, the object and subject of the research are defined, and the scientific novelty, theoretical, and practical significance of the work are revealed.

The first section of this dissertation analyzes the current state of scheduling and data mining methods in manufacturing systems. Mathematical formulations of the Job Shop Scheduling Problem (JSSP) are considered. Shop Scheduling Problem (JSSP), discrete optimization methods, and robust planning approaches are explored. The potential of machine learning methods for predicting production process parameters and assessing equipment performance is analyzed. The integration of predictive models and optimization algorithms is shown to be a promising direction for the development of intelligent planning systems.

In the second section, a methodological framework for a hybrid predictive-optimization approach to scheduling is developed. The scheduling problem is

formalized, taking into account the uncertainty of operation durations and equipment degradation risks. An extended optimization model for scheduling is proposed, using quantile forecasts of operation durations and aggregated PdM risk indicators. A scenario-based mechanism for applying predictive constraints and an adaptive scheme for relaxing them in the absence of feasible solutions are developed.

In the third section, machine learning models for predicting operation durations and assessing equipment degradation risks were developed. A hybrid quantile forecasting model based on ensemble and neural network methods was implemented, enabling the formation of P10, P50, and P90 estimates of operation durations. A predictive equipment maintenance model based on diagnostic data was developed. Computational experiments were conducted, confirming the feasibility of using predictive estimates as parameters in an optimization scheduling model.

In the fourth section, an experimental study of the proposed hybrid ML-CP-SAT approach is conducted using benchmark instances of the scheduling problem. Various scheduling scenarios are compared using forecast constraints and PdM risks. The impact of the scenarios on the objective function value, the stability of solutions, and the computational characteristics of the optimization process is analyzed. It is shown that the use of forecast models improves the controllability of the optimization search and the stability of the scheduling process.

The fifth section presents the practical implementation of an intelligent scheduling system as a software prototype integrating machine learning modules, PdM risk assessment, and the CP-SAT optimization solver. The system's user interface was developed, single-instance and batch scheduling modes were implemented, and a decision support module was added. Experimental testing of the developed prototype confirmed the viability of the proposed approach and its applicability to production planning and educational research.

The conclusion presents the main results of the study and formulates conclusions confirming the achievement of the stated goal and the solution of the tasks of the dissertation. The appendices contain copyright certificates and an implementation report.

The dissertation research was carried out within the framework of the program-targeted funding project BR24992785 «Organization and implementation of comprehensive research to ensure the sustainable development of the agro-industrial complex of the Kostanay region with the establishment of a research and technological center», implemented in accordance with the priority directions of the scientific and technological development of the Republic of Kazakhstan.