Syllabus
«Intelligent technologies of analysis and data pre-processing»

Lecturer: *Irina Kalinina*
PhD, Associate Professor, Associate Professor of the Intelligent Information Systems Department

Expected learning outcomes
As a result of studying the discipline, the student
must know:

- conceptual and methodological knowledge in the field of data analysis for optimal decision-making, as well as research skills sufficient for conducting scientific and applied research in the field of data analysis at the level of the latest world achievements in computer engineering, IT infrastructures, information technologies;
- modern methods of conducting research in the field of applied data analysis for decision-making in various fields of science by studying the theoretical and practical provisions of building behavioural models of functional dependencies and using the results of data analysis to clarify the optimal decision-making parameters;
- scientific and mathematical provisions underlying data analysis methods for obtaining optimal decision-making parameters, methods of building and researching mathematical models and adaptive and intelligent computing technologies in data analysis.

***must be able to*:**

- effectively search and critically analyse information from various sources regarding data analysis methods and technologies;
- solve problems of synthesis and analysis of research objects when analysing data to make optimal decisions;
- create and implement behavioural mathematical models for describing functional dependencies in data analysis;

The scope of the discipline: 5,5 ECTS credits (16 hours of lectures, 48 hours of practice).

Purpose: training of specialists capable of solving complex problems in the field of research activities in the field of data analysis for making optimal decisions and using the results of data analysis to clarify scientific conclusions and form forecasts regarding the decisions made, which involves a deep understanding of approaches to creating traditional and creating new ones behavioural models.

The content of the discipline

Topic 1. General concept of data analysis and scope of applied analysis tasks. Comparison of machine learning and intelligent data analysis methods. Classification of machine learning problems. Problems of probabilistic statistical analysis and data pre-processing in machine learning problems.
Topic 2. General concept of applied problems of data analysis. Classification tasks. The task of evaluating borrowers. Problems of regression estimation. The task of forecasting consumer demand. Problems of clustering. The task of classifying texts. The task of selecting terms. The task of a sociological survey. Association search tasks**.**
Topic 3. Statement of the problem of data analysis. Basic definitions. Objects and signs. Types of data analysis tasks. Concept of data analysis algorithm. Definition of the teaching method. Definition and methods of formation of the loss function. Functionality of quality. Minimization of empirical risk. Cross-validation. Leave-one-out cross-validation.
Topic 4. Methods of identification and filling gaps in data**.**
Topic 5. Methods and algorithms for detecting and processing extreme values. Emission identification methods.
Topic 6. Data smoothing methods and algorithms**.**
Topic 7. Systematic use of data filtering methods in machine learning tasks. Digital filtering. Optimal filtering. Probabilistic filtering**.**
Topic 8. A systematic approach to data normalization and standardization in machine

- apply applied methods of data analysis of adaptive and intelligent computing technology to predict future states of researched objects and processes.
- apply application libraries and software systems used in machine data analysis;
- to have programming methods and technologies using application libraries and software systems designed for machine analysis of real data sets.

Prerequisites
According to the educational program, it is necessary to acquire knowledge in the following disciplines: "Higher mathematics", "Fundamentals of programming", "Object-oriented programming", "Probability theory and mathematical statistics", "Algorithms and data structures", "System analysis", "Methods and systems of artificial intelligence"**.**

Consequences
Competences, knowledge and skills acquired within the framework of studying this discipline can be applied to obtain reasonable research results and increase the scientific level of qualification work**.**

Technical support
Laboratory work on the discipline is conducted in computer classes using languages and software environments: R Studio, R, Python**.**

Deadline Policy
Works that are submitted late without good reason will be assigned a lower grade**.**

Academic Integrity Policy
Provides independent performance of practices. Write-off (including using mobile devices) is prohibited. The work is not credited in case of detection of plagiarism or write-off.

learning tasks. Linear methods of normalization. Nonlinear methods of normalization**.**
Topic 9. Information technology of probabilistic statistical analysis and data pre-processing.

Evaluation criteria of laboratory works / practices / individual works / reports / projects

Maximum number of points a student with high quality independently completed the entire amount of work, answers all questions related to the completed work, and makes additional calculations offered by the teacher. The teacher has no complaints about the implementation and requirements for the performance of the work.
Approximately 70%-99% of the maximum number of points – a student completed all tasks with sufficient quality, but in the process of work he made some mistakes, which, after being pointed out by the teacher, he corrected himself. He answers some questions incorrectly. Additional calculations proposed by the teacher are done with some effort. Not all requirements for performance of work are met.
Approximately 40%-69% of the maximum number of points – a student independently completed all the work, but the quality of the implementation is insufficient (calculation errors, not all work requirements are met). The answers to questions about the performance of work are not quite clear. There are mistakes in the answers.
Approximately 1%-39% of the maximum number of points – a student did not complete all the work independently, while the quality of the implementation was insufficient (errors in calculations, does not comply with the requirements for the design of the work). He does not answer questions about the performance of work clearly. There are gross mistakes in the answers.
0 points – a student did not complete the entire amount of work, or did it with gross errors. He has problems with calculations, does not know the theoretical material, the software implementation does not meet the requirements.