Syllabus
Course Program

DETECTION OF ANOMALIES IN DATA AND TIME SERIES

Specialty
113 Applied mathematics

Institute
Educational and Scientific Institute of Computer Sciences and Information Technologies

Educational program
Intelligent data analysis

Department
Computer Mathematics and Data Analysis

Level of education
Bachelor's level

Course type
Special (professional), Selective

Semester
8

Language of instruction
Ukrainian

Lecturers and course developers

First name and surname

Sergey Garder Sergei.Garder@khpi.edu.ua
Candidate of technical sciences, associate professor of the department of computer mathematics and data analysis of NTU "KhPI".
Work experience - 32 years. Author of 89 scientific and educational and methodological works. Leading lecturer in the disciplines: "Data analysis", "Theory of time series", "Artificial neural networks"
Learn more about the teacher on the department's website

General information

Summary

The purpose of teaching the discipline is to provide training for specialists in the field of applied mathematics and cyber security capable of formulating, solving and generalizing practical problems in their professional activities using fundamental and special applied methods of data analysis, time series theories, neural network technologies. To develop mathematical models, algorithms, create and operate the relevant software...

Course objectives and goals

The purpose of teaching the discipline is to provide training for specialists in the field of applied mathematics and cyber security capable of formulating, solving and generalizing practical problems in their professional activities using fundamental and special applied methods of data analysis, time series theories, neural network technologies. To develop mathematical models, algorithms, create and operate the relevant software.ls

Format of classes

Lectures, laboratory classes, consultations, self-study. Final control in the form of an exam.

Competencies

GC3: Ability for continuous learning, acquiring new knowledge and skills, including in areas other than professional ones.
GC4: Ability to identify, pose, and solve problems in professional activities.

GC5: Ability to generate new ideas (creativity) and unconventional approaches to their implementation, flexible adaptation to real professional situations, displaying a creative approach and initiative.
GC6: Ability to critically evaluate and rethink accumulated experience (own and others'), analyze one's professional and social activities.
GC7: Ability to work with information, find and use information from various sources necessary for solving professional tasks.
GC8: Effective communication skills, taking into account the goals and situation of communication.
SC1: Ability to formulate a mathematical problem, relying on the language of the subject area, verifying the correctness of the formulation, including under conditions of uncertainty.
SC2: Ability to choose, develop, and investigate mathematical, analytical, or numerical methods for solving practical problems that ensure the required accuracy and reliability of the result.
SC3: Ability to choose, develop, investigate, and apply mathematical methods for solving practical problems of modeling, design, management, forecasting, decision-making.
SC10: Ability to choose, develop, investigate, and apply mathematical models and methods for intelligent data analysis under conditions of uncertainty.

Learning outcomes

LO2: Ability to formalize problems formulated in the language of a specific subject area, choose a rational method for solving them, solve problems using analytical or numerical methods, assess the accuracy and reliability of the results, and interpret them.
LO4: Ability to combine methods of mathematical and computer modeling with informal expert analysis procedures to search for optimal solutions.
LO7: Apply modern programming technologies and software development, implement numerical and symbolic algorithms.

Student workload

The total volume of the discipline is 120 hours. (4 ECTS credits): lectures – 20 hours, laboratory work – 20 hours, independent work – 80 hours.

Course prerequisites

Algorithmization and programming", "Numerical methods", "Analysis of data and time series"..

Features of the course, teaching and learning methods, and technologies

When teaching the discipline "Neural Network Technologies", active and interactive methods of teaching lectures, laboratory classes on real data in an active form, and collective discussion of problems are provided. The effectiveness of the educational process is manifested in the increase of self-awareness of students; formation of the ability to make independent decisions, acquisition of skills for collective discussion of problems; development of analytical and logical abilities. Study materials are available to students through OneNote Class Notebook, Teams..

Program of the course

Topics of the lectures

Topic 1. Introduction to the concept of time series. Time series as a discrete random process Genesis of time series data, tasks and goals of time series research. Time series smoothing methods. Additive time series model. Triple smoothing algorithm, S(ARIMA), boosting. Analytical methods of exclusion of the deterministic component. . Identification of the trend of time series by analytical methods and analysis of a series of residuals. Trend spline analysis

Topic 2. The task of detecting anomalies in data.

Fields of application. Forecasting of equipment breakdowns, Forecasting of fraudulent activities, detection of abnormal consumption patterns, Detection of abnormal demand and load Primary time series analysis. Irwin's criterion for finding abnormal values.

Topic 3. Types of anomalies

The task of detecting anomalies in data

point anomalies, when there is a deviation in behavior at individual points;

group anomalies, in which a group of points behaves abnormally, context anomalies, when the anomaly is related to external data, not related to the values of the series (for example, negative temperature outside in summer).
Topic 4 Methods of detecting anomalies.
Method of support vectors with one class One-Class SVM Method of isolating forest - isolate forest Method "Elliptic envelope" and statistical methods.
Topic 5 Methods of detecting anomalies.
Metric methods (algorithms "k nearest neighbors", "k-th nearest neighbor", ABOD (angle-based outlier detection) or LOF (local outlier factor)). Cluster methods.
Topic 5 (continued). Methods of detecting anomalies. Method of principal components.
Topic 6 Application of recurrent neural networks
Topic 5. (continued). Regression methods for detecting anomalies....

Topics of the workshops

Practical classes within the discipline are not provided.

Topics of the laboratory classes

1. Time series smoothing methods. Initial check for the presence of abnormal values and their exclusion.
2. Analytical methods of exclusion of the deterministic component. . Trend identification
3. Method of support vectors with one class One-Class SVM
4. "k nearest neighbors" and "k-th nearest neighbor" algorithms.
5. Method of principal components.
6. Regression methods for detecting anomalies.
7. Clustering for finding anomalies.
8. Clustering to find anomalous values (using Kohonen ANN, Kohonen map).
9. Modeling time series using Elman's recurrent network.

Self-study

The course involves the implementation of an individual calculation task on modeling (using PITON) a real time series or a data set of the student's choice. The result of calculations and modeling is drawn up in a written report. Students are also recommended additional materials (articles [9], [10], [11]) for independent study and analysis.
Non-formal education
Within the framework of non-formal education according to the relevant Regulation (http://surl.li/pxssv), the educational component or its separate topics can be taken into account in case of independent completion of professional courses/trainings, obtaining civic education, online education, professional internship, etc.
In particular, individual topics of this component may be taken into account upon successful completion of the following courses:
- Topic.4 "Research of anomalies"
https://prometheus.org.ua/course/course-v1:IRF+Stat101+2016

Course materials and recommended reading

Basic literature

Assessment and grading

Criteria for assessment of student performance, and the final score structure

Description of the final score structure, course requirements, and necessary steps to earn points, especially paying attention to self-study and individual assignments.

Grading scale

Total points	National	ECTS
90–100	Excellent	A
82–89	Good	B
75–81	Good	C
64–74	Satisfactory	D
60–63	Satisfactory	E
35–59	Unsatisfactory (requires additional learning)	FX
1–34	Unsatisfactory (requires repetition of the course)	F

Norms of academic integrity and course policy

The student must adhere to the Code of Ethics of Academic Relations and Integrity of NTU "KhPI": to demonstrate discipline, good manners, kindness, honesty, and responsibility. Conflict situations should be openly discussed in academic groups with a lecturer, and if it is impossible to resolve the conflict, they should be brought to the attention of the Institute's management.
Regulatory and legal documents related to the implementation of the principles of academic integrity at NTU "KhPI" are available on the website: http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/

Approval

Approved by	Date, signature 29.08.2024	Head of the department Olena AKHIIEZER
	Date, signature 29.08.2024	Guarantor of the educational program Olena AKHIIEZER