

INTELLIGENT DATA ANALYSIS

СИЛАБУС

Code and name of specialty	122 Computer Science	Institute / faculty	Faculty of Computer Science and Software Engineering
Program name	"Computer Science and Intelligent Systems"	Department	Software Engineering and Management Information Technologies
Type of program	Educational and Professional	Language of instruction	Ukrainian, English

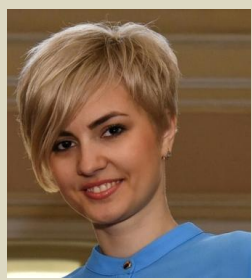
Lecturer

Full name, e-mail

Olha Yanholenko,
Olha.Yanholenko@khpi.edu.ua
Anna Goloskokova
anna.goloskokova@khpi.edu.ua



PhD, Candidate of Engineering Sciences (05.13.06 – information technologies), Associate Professor of Department of Software Engineering and Management Information Technologies. Work experience – since 2012. Author (co-author) of more than 50 scientific and educational publications (h-index= 4, i10-index= 1 in Google Scholar - <https://scholar.google.com/citations?hl=en&user=QBNVtqEAAAAJ>; ORCID <https://orcid.org/0000-0001-7755-1255>). Main courses: «Intelligent data analysis» (lectures and laboratory classes), «Intelligent data analysis and knowledge retrieval» (lectures and laboratory classes in English and Ukrainian).



PhD, Candidate of Engineering Sciences (05.13.06 – information technologies), Associate Professor of Department of Software Engineering and Management Information Technologies. Work experience – since 2011. Prepared and published more than 20 publications, 1 collective monographs, 2 articles in publications indexed in Scopus and Web of Science.
<https://scholar.google.com.ua/citations?hl=uk&user=jnDzQRAAAAAJ>
<https://orcid.org/0000-0001-9012-7889>
<https://www.scopus.com/authid/detail.uri?origin=resultslist&authorId=57190428440&zone=57190428440>
Extensive teaching at the university level: Bachelor: Artificial intelligence systems (in English), Introduction into neural networks (in English and Ukrainian), Basics of soft computing theory (in English and in Ukrainian)

GENERAL DESCRIPTION OF THE COURSE

Summary

The course "Intelligent Data Analysis" is a discipline in the cycle of selective training in the specialty 122 "Computer Science". It is taught in the sixth semester in the amount of 90 hours (3 ECTS credits), in particular: lectures - 20 hours, laboratory classes - 10 hours, independent work - 60 hours. There are no individual tasks. The study of the discipline ends with a test.

Course objectives

Formation of students' theoretical and practical knowledge on intelligent data analysis in the process of software systems design and implementation, application of data mining techniques to software design problem solving.

Types of classes and control	Lectures, laboratory classes. Current control - laboratory work, intermediate modular control. Final control – credit.						
Term	8						
Student workload (credits) / Type of course (mandatory / elective)	3/ mandatory	Lectures (hours)	20	Workshops (hours)	10	Self-study (hours)	60
Program competences	<p>GC1. Ability to abstract thinking, analysis and synthesis.</p> <p>GC2. Ability to apply knowledge in practical situations.</p> <p>GC3. Knowledge and understanding of the subject area and understanding of professional activity.</p> <p>GC6. Ability to learn and master modern knowledge.</p> <p>GC7. Ability to search, process and analyze information from various sources.</p> <p>PC2. Ability to detect statistical patterns of non-deterministic phenomena, the use of computational intelligence methods, including statistical, neural network and fuzzy data processing, machine learning and genetic programming methods, etc.</p> <p>PC7. Ability to apply the theoretical and practical basics of methodology and modelling technology to study the characteristics and behavior of complex objects and systems, to conduct computational experiments with processing and analysis of results.</p> <p>PC11. Ability to conduct intelligent data analysis based on methods of computational intelligence, including large and poorly structured data, their operational processing and visualization of analysis results in the process of solving applied problems.</p> <p>PC16. Ability to implement high-performance computing based on cloud services and technologies, parallel and distributed computing in the development and maintenance of distributed parallel information processing systems.</p> <p>PC17. Ability to apply the theoretical and practical basics of modern management theory for complex organizational, technical and socio-economic systems to build intelligent management systems, in the process of designing intelligent systems to use modern information processing technologies and methods of computational intelligence.</p>						
Learning outcomes	Teaching and learning methods			Forms of assessment (continuous assessment CAS, final assessment FAS)			
PLO3. Use knowledge of the laws of random phenomena, their properties and operations with them, models of random processes, and modern software environments to solve problems of statistical data processing and construction of predictive models.	In the process of teaching is used such initial technologies as: lectures, laboratory work, presentations that develop communication and leadership skills, independent work with literary sources, mixed forms of learning using distance platforms			Current CAS assessment: Assessment of students' work in the laboratory Intermediate modular control Final FAS assessment: Credit			
PLO4. Use methods of computational intelligence, machine learning, neural network, and fuzzy data processing, genetic and evolutionary programming to solve problems of recognition, prediction, classification, identification of management objects, etc.	In the process of teaching is used such initial technologies as: lectures, laboratory work, presentations that develop communication and leadership skills, independent work with literary sources, mixed forms of learning using distance platforms			Current CAS assessment: Assessment of students' work in the laboratory Intermediate modular control Final FAS assessment: Credit			
PLO8. Use the methodology of system analysis of objects, processes, and systems for the tasks of analysis, prediction, management, and design of dynamic processes	In the process of teaching is used such initial technologies as: lectures, laboratory work, presentations that develop communication and			Current CAS assessment: Assessment of students' work in the laboratory Intermediate modular control			

in macroeconomic, technical, technological, and financial objects.	leadership skills, independent work with literary sources, mixed forms of learning using distance platforms	Final FAS assessment: Credit
PLO12. Apply methods and algorithms of computational intelligence and intelligent data analysis in the tasks of classification, prediction, cluster analysis, search for associative rules using software tools to support multidimensional data analysis based on technologies DataMining, TextMining, WebMining.	In the process of teaching is used such initial technologies as: lectures, laboratory work, presentations that develop communication and leadership skills, independent work with literary sources, mixed forms of learning using distance platforms	Current CAS assessment: Assessment of students' work in the laboratory Intermediate modular control Final FAS assessment: Credit
PLO16. Perform parallel and distributed computations, apply numerical methods and algorithms for parallel structures, parallel programming languages in the development and operation of parallel and distributed software.	In the process of teaching is used such initial technologies as: lectures, laboratory work, presentations that develop communication and leadership skills, independent work with literary sources, mixed forms of learning using distance platforms	Current CAS assessment: Assessment of students' work in the laboratory Intermediate modular control Final FAS assessment: Credit
PLO17. Apply for the construction of intelligent management systems theoretical and practical foundations of modern management theory, design intelligent systems using modern information processing technologies and methods of computational intelligence.	In the process of teaching is used such initial technologies as: lectures, laboratory work, presentations that develop communication and leadership skills, independent work with literary sources, mixed forms of learning using distance platforms	Current CAS assessment: Assessment of students' work in the laboratory Intermediate modular control Final FAS assessment: Credit

ASSESSMENT AND GRADING

Ranges of points corresponding to grades	Total score (points) for all types of learning activities	ECTS grading scale	The national grading scale	Allocation of grade points
	90-100	A	Excellent	
	82-89	B	Good	
	74-81	C		
	64-73	D	Satisfactory	
	60-63	E		
	35-59	FX	Unsatisfactory (with the exam retake option)	
0-34	F	Unsatisfactory (with mandatory repetition of the course)		

100% final assessment in the form of credit **(30%)** and current assessment **(70%)**.
30% credit
70% current rating:
 Module №1 (10%)
 Module №2 (20%)
 Laboratory work (40%)

Course policy	The student is required to attend all classes according to the curriculum and adhere to the norms of academic ethics. To study the discipline you need to have your own personal computer and / or use the computers of the computer center of the department. The student must work with required and
----------------------	--

additional literature, including information resources on the Internet. All laboratory work must be completed and submitted by the student during the semester in which the discipline is taught, before the start of the test week. Without the personal presence of the student the final control is not carried out.

COURSE STRUCTURE AND CONTENT

Lecture 1	Introduction to intelligent data analysis. Data mining definition, needs and purposes.	Laboratory work 1	Data preprocessing with Weka.	Individual work	Regression problems.
Lecture 2	Data mining process. CRISP-DM process. SEMMA process. Data mining process applications	Laboratory work 2	Clustering problem solving with Weka.		Data analysis visualization.
Lecture 3	Data pretreatment. Handling noninformative variables, missing data, data cleaning.	Laboratory work 3	JSON data processing using Python.		Text mining algorithms. Stemming, stop words elimination.
Lecture 4	Sampling schemes for fixed datasets. Reducing number of attributes.	Laboratory work 4	XML data processing using Python.		Web mining algorithms. Web documents retrieval and analysis.
Lecture 5	Memory-based reasoning methods. Matching, weighted matching.	Laboratory work 5	Web crawler implementation using Python.		Agent systems for web mining algorithms implementation.
Lecture 6	Association rules in knowledge discovery. Market basket analysis.				
Lecture 7	Clustering. K-means clustering.				
Lecture 8	Text mining. Representing text documents. Constructing a vector space model.				
Lecture 9	Web mining. Web content, web usage, web structure mining.				
Lecture 10	Intelligent methods of web data processing.				

RECOMMENDED READING

Compulsory	<ol style="list-style-type: none"> 1. Kretowski M. (2019) (Evolutionary Decision Trees in Large-Scale Data Mining. 2. Ranga Suri N.N.R., Murty N., Athithan M.G. (2019) Outlier Detection: Techniques and Applications. A Data Mining Perspective. -Cham: Springer International Publishing. 3. Anandan R. (ed.) (2021) A Closer Look at Big Data Analytics. - Nova Science Publishers, Inc. 4. Samanta D., Banerjee A. (2021) Computationally Intensive Statistics for Intelligent IoT. Springer. 5. Akerkar R. (Ed.) (2020) Big Data in Emergency Management: Exploitation Techniques for Social and Mobile Data. - Springer 6. Moreira, J. (2018) Data Analysis Made Easy. New York: Wiley-Interscience. 7. Srinivas M., Sucharitha G, Matta A., Chatterjee P. (eds.) (2021) Machine Learning Algorithms and Applications: Theory and Applications. Wiley-Scrivener Publishing, 	Recommended	<ol style="list-style-type: none"> 1. Scitovski R., Sabo K., Martínez-Álvarez F., Ungar S. (2021) Cluster Analysis and Applications. Springer, — 2. Бондаренко М. Ф., Шабанов-Кушнарченко Ю. П., Сергиенко И. В.(2011) Мозгоподобные структуры: Справочное пособие. Том первый. Киев: Наукова думка. 3. Tyagi A.K. (2021) Data Science and Data Analytics: Opportunities and Challenges. New York: Chapman and Hall/CRC. 4. Sakarkar G., Patil G., Dutta P.(2021) Machine Learning Algorithms Using Python Programming. Nova Science Publishers, Inc. 5. Wooldridge M. J. (2009) An introduction to multiagent systems. 6. Бондаренко М. Ф., Шабанов-Кушнарченко Ю. П. Теория интеллекта: Харьков: ООО «Компания СМИТ». <p>ІНФОРМАЦІЙНІ РЕСУРСИ В ІНТЕРНЕТІ</p> <ol style="list-style-type: none"> 1. Weka 3: Machine Learning Software in Java [Електронний ресурс] – Режим доступу: https://www.cs.waikato.ac.nz/ml/weka. 2. Data Mining: Practical Machine Learning Tools and Techniques[Електронний ресурс] – Режим доступу: https://www.cs.waikato.ac.nz/ml/weka/book.html.
Academic integrity			
The student must adhere to the Code of Ethics of Academic Relations and Integrity of NTU "KhPI": to show discipline, politeness, friendliness, honesty, responsibility			
The content of this syllabus is consistent with the course program.			