Intelligent data analysis

Specialty	**Institute**
121 – Software Engineering	Institute of Computer Science and Information Technology
Educational program	**Department**
Software Engineering	Software Engineering and Management Intelligent Technologies (321)
Level of education	**Course type**
Bachelor's level	Special (professional), Mandatory
Semester	**Language of instruction**
8	English, Ukrainian

Lecturers and course developers

Svitlana Yershova
Svetlana.Ershova@khpi.edu.ua
Senior Lecturer of Department of Software Engineering Management Intelligent technologies, National Technical University "Kharkiv Polytechnic Institute"
Prepared and published more than 20 publications.
Developed and taught more than 15 different training courses
Google Scholar: https://scholar.google.com.tw/;
ORCID: https://orcid.org/0000-0003-3893-117X
Scopus (Web of Science): https://ceur-ws.org/Vol-2753/paper25.pdf
https://www.scopus.com/record/display.uri?eid=2-s2.0-85134395161&origin=resultslist&sort=plf-f&retries=1,
https://ceur-ws.org/Vol-3403/paper37.pdf.

More about the lecturer on the department's website

General information

Summary

The discipline program focuses on the consideration of technologies, basic concepts, principles, modern achievements and prospects for the development of intelligent data analysis. The processes of data mining, data pre-processing, basic concepts of the theory of associative rules, methods of cluster analysis are considered; the main issues related to building a vector space model and conducting multivariate data analysis based on Data Mining, Text Mining, Web Mining technologies. The acquisition of practical skills of intelligent data analysisis carried out on the basis of Python, WEKA.

Course objectives and goals

The purpose and task of the discipline " Intelligent data analysis" is to form a complex of knowledge among computer science specialists necessary to understand the problems that arise during intelligent data analysis, mastering models and methods for solving data mining problems. The training course occupies a leading place in the formation of the future specialist in computer science. The knowledge and skills obtained in the process of studying the discipline " Intelligent data analysis" are integral components of the formation of professional competence and an important aspect of the professional training of students.

Format of classes

The discipline " Intelligent data analysis" is an academic discipline from the cycle of special (professional) training in specialty 122 - "Computer Science. "
Lectures, laboratory works, independent work, calculation task, consultations. Final control – standings.

Competencies

GC1. Ability to think abstractly, analyze and synthesize.
GC2. Ability to apply knowledge in practical situations.
GC3. Knowledge and understanding of the subject area and understanding of professional activities.
GC6. Ability to learn and master modern knowledge.
GC7. Ability to search, process and analyze information from various sources.
PC2. Ability to identify statistical regularities of non-deterministic phenomena, apply methods of computational intelligence, in particular statistical, neural network and fuzzy data processing, machine learning and genetic programming methods, etc.
PC11. Ability to intelligently analyze data based on computational intelligence methods, including large and poorly structured data, their operational processing and visualization of analysis results in the process of solving applied problems.
PC16. Ability to implement high-performance computing based on cloud services and technologies, parallel and distributed computing in the development and operation of distributed parallel information processing systems.
PC17. Ability to apply the theoretical and practical foundations of modern management theory of complex organizational, technical and socio-economic systems to build intelligent management systems, to use modern information processing technologies and methods of computational intelligence in the design of intelligent systems.

Learning outcomes

 PLO3. To use knowledge of the laws of random phenomena, their properties and operations on them, models of random processes and modern software environments to solve problems of statistical data processing and build predictive models.
PLO4. To use methods of computational intelligence, machine learning, neural network and fuzzy data processing, genetic and evolutionary programming to solve problems of recognition, forecasting, classification, identification of control objects, etc.
PLO12. Apply methods and algorithms of computational intelligence and data mining in the tasks of classification, forecasting, cluster analysis, search for associative rules using software tools to support multivariate data analysis based on Data Mining, Text Mining, Web Mining technologies.
PLO16. Perform parallel and distributed computing, apply numerical methods and algorithms for parallel structures, parallel programming languages in the development and operation of parallel and distributed software.
PLO17. Apply theoretical and practical foundations of modern management theory to build intelligent control systems, design intelligent systems using modern information processing technologies and methods of computational intelligence.

Student workload

The discipline is taught in the 8th semester in the amount of 120 hours. (4 ECTS credits), in particular: lectures – 20 hours, laboratory – 20 hours, calculation task, independent work – 80 hours.

Course prerequisites

The course " Intelligent data analysis" is based on the disciplines "Discrete Mathematics", "Probability Theory and Mathematical Statistics", "Databases", "Cybersecurity"

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

Teaching and learning methods:

National Technical University
"Kharkiv Polytechnic Institute"

interactive lectures with presentations, discussions, laboratory classes, teamwork, case method, student feedback, problem-based learning.

Forms of assessment:
written individual assignments for laboratory work (CAS), assessment of knowledge in laboratory classes (CAS), express surveys (CAS), online tests (CAS), final/semester control in the form of a semester exam, according to the schedule of the educational process (FAS).

Program of the course

Topics of the lectures

Topic 1. "Basic concepts of data mining".
The concept of "information, data, knowledge." The essence of data mining. Objectives of data mining. Areas of practical use of data mining.
Topic 2. «Data mining processes".
General information about the data mining process CRISP-DM process. SEMMA process. Application of data mining processes.
Topic 3. "Data preprocessing".
Data pre-processing. Processing of uninformative variables, missing data, data cleaning.
Topic 4. "Reducing the number of attributes."
Sampling schemes for fixed data sets. Reduce the number of attributes.
Topic 5. "Reasoning based on memory."
Methods of reasoning based on memory. Coincidence, weighted coincidence.
Topic 6. "Associations".
Basic concepts of the theory of associative rules. Association rules in the extraction of knowledge. Consistent association. Algorithms of the family "Apriori".
Topic 7. "Clustering".
The essence of clustering. Data preparation. Algorithm for cluster analysis. Statistics related to cluster analysis. Model testing. Cluster analysis methods.
Topic 8. "Text Mining".
Text extraction. Presentation of text documents. Construction of a vector space model.
Topic 9. " Web Mining".
 Web Mining. Web content, internet usage, web structures mining.
Topic 10. " Web Data Processing".
Intelligent web data processing techniques.

Topics of the workshops

Workshops are not provided within the discipline.

Topics of the laboratory classes

Topic 1.
 Pre-processing data using Weka.
Topic 2.
Solving the problem of clustering using Weka.
Topic 3.
Processing JSON data using Python.
Topic 4.
Processing XML data using Python.
Topic 5.
Implementing a web crawler with Python.

Self-study

Students are recommended additional materials (videos, articles) for self-study and processing.

National Technical University "Kharkiv Polytechnic Institute"

Course materials and recommended reading

Key literature

1. Берри Пол Head First Python. Легкий для сприйняття довідник/ Пол Берри, - Видавництво Фабула, 2021, 624 с.
2. Іванов С.М., Максишко Н.К., Бречко Д.О. Інтелектуальний аналіз даних: конспект лекцій/ С.М. Іванов, Н.К. Максишко Н.К., Д.О. Бречко,- Запоріжжя: ЗНУ, 2020, 156 с.
3. Kretowski, M. Evolutionary Decision Trees in Large-Scale Data Mining /M. Kretowski, - Cham: Springer International Publishing, 2019, 180 p.
4. Ranga Suri N.N.R., Murty N., Athithan M.G. Outlier Detection: Techniques and Applications. A Data Mining Perspective / N.N.R. Ranga Suri, N. Murty, M.G. Athithan, - Cham: Springer International Publishing, 2019, 214 p.
5. Samanta D., Banerjee A. Computationally Intensive Statistics for Intelligent/ D. Samanta, A.Banerjee ,- Springer, 2021, 218 p.

Additional literature

1. Akerkar Ranjendra Big Data in Emergency Management: Exploitation Techniques for Social and Mobile Data/ Ranjendra Akerkar, - Springer, 2020, 201 p.
2. Anandan, R. A Closer Look at Big Data Analytics/ R. Anandan, - Nova Science Publishers. Inc., 2021, 366 p.
3. Botros Silvia, Tinley Jeremy High Performance MySQL: Proven Strategies for Operating at Scale 4th Edition/ Silvia Botros, Jeremy Tinley, - O'Reilly Media, 2021, 388 p.
 4. Sakarkar Gaurav, Patil Gaurav, Dutta Preteek Machine Learning Algorithms Using Python Programming/ Gaurav Sakarkar, Gaurav Patil, Preteek Dutta, - Nova Science Publishers. Inc., 2021, 218 p.
5. Srinivas M., Sucharitha G., Matta A., Chatterjee P. /M. Srinivas, G. Sucharitha, A. Matta, P. Chatterjee Machine Learning Algorithms and Applications: Theory and Applications, -Wiley-Scrivener Publishing, 2021, 368 p.

Assessment and grading

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

100% of the final grade consists of the results of the assessment in the form of an exam (40%) and current assessment (60%):
- 5 laboratory works (5% each);
- 1 test paper (5%);
- 1 calculation task (10%);
- independent work (20%).

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/

National Technical University "Kharkiv Polytechnic Institute"

Approval

Approved by   11.04.2023