COURSE DESCRIPTION CARD - SYLLABUS

Course name
Data Compression and Big Data Methods [S2Teleinf2-STRC>KD]

Course

Field of study
Teleinformatics

Year/Semester
2/3

Area of study (specialization)
ICT networks and cloud solutions

Profile of study
general academic

Level of study
second-cycle

Course offered in
Polish

Form of study
full-time

Requirements
compulsory

Number of hours

Lecture
14

Laboratory classes
24

Other (e.g. online)
0

Tutorials
0

Projects/seminars
0

Number of credit points

4,00

Coordinators

dr hab. inż. Damian Karwowski
damian.karwowski@put.poznan.pl

Lecturers

Prerequisites

1. The student has systematic knowledge of mathematical analysis, algebra and probability theory. 2. Has structured and mathematically based knowledge of the theory of one-dimensional signals necessary to understand the representation and analysis of signals in the time domain and in the frequency domain. 3. Student knows the principles of constructing computer programs, has knowledge in the field of computer science and knows the syntax of selected high-level programming languages (e.g.: C, C++, C#, Python, MatLab).

Course objective

The aim of the course is to present selected advanced data compression techniques (that implement complex mechanisms of data prediction), methods of data analysis and methods of statistical inference. Moreover, the aim of the course is to present selected techniques of exploratory data analysis, including methods of classifying and grouping large data sets.

Course-related learning outcomes

Knowledge:
1. The student has structured, mathematically based knowledge of advanced data compression

techniques. In addition, the student has knowledge of exploratory data analysis methods, including methods of data classification and grouping (K2_W01, K2_W02, K2_W03).
2. Has knowledge of the principles of operation of the known data compression and mining algorithms and is able to apply the known methods for efficient data representation and analysis (K2_W01, K2_W02, K2_W03).
3. Knows the advantages and limitations of the known methods, understands the benefits of using the known solutions for the purposes of efficient data representation and analysis (K2_W07).

Skills:
1. Student is able to present the algorithm and mathematical description of the known data compression and analysis methods and to propose the appropriate method of data coding and analysis (K2_U06).
2. Is able to compress data in order to represent it efficiently and analyze data in order to obtain selected parameters of the data set (K2_U07).
3. Using the acquired methods, he can design his own data compression and analysis method (K2_U09, K2_U10).

Social competences:
Student understands the need for continuous education in order to improve professional qualifications (K2_K01, K2_K06).

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

1. Lecture
Written and/or oral examination. The colloquium consists of several or a dozen questions (depending on the nature of the questions) and concerns the content presented during the lectures. The exact nature of the exam questions will be presented to students during one of the last lectures. The threshold for passing the exam: 50% of points.
Grading scale: <50% - 2.0 (ndst); 50% to 59% - 3.0 (dst); 60% to 69% - 3.5 (dst+) ; 70% to 79% - 4.0 (db); 80% to 89% - 4.5 (db+); 90% to 100% - 5.0 (bdb).
2. Laboratories
Colloquium at the end of the semester and/or tests to check the level of mastery of current material, and/or assessment of reports prepared by students. The colloquium/tests consist of several/dozen examination questions, depending on the nature of the questions asked. The exact nature of the questions will be presented to students before the colloquium/test. Passing threshold: 50% of points.
Grading scale: <50% - 2.0 (ndst); 50% to 59% - 3.0 (dst); 60% to 69% - 3.5 (dst+) ; 70% to 79% - 4.0 (db); 80% to 89% - 4.5 (db+); 90% to 100% - 5.0 (bdb).

Programme content

1. Advanced data compression methods.
2. Exploratory data analysis.
3. Processing large data sets using machine learning techniques and artificial neural networks.

Course topics

1. Lecture
Advanced data compression, including complex data prediction mechanisms, prediction residual signal analysis methods, data statistics estimation algorithms.
Analysis and assessment of the efficient of modern data compression methods (based on the latest available solutions), and assessment of the computational complexity of advanced compression methods.
Exploratory data analysis (including descriptive statistics of data). Examples of data mining techniques, including data classification and grouping methods and regression mechanism.
Processing large data sets using machine learning techniques and artificial neural networks.
2. Laboratories
Data prediction techniques (measuring the efficiency of selected methods).
Methods of encoding the residual data prediction signal (measuring the efficiency of selected methods).
Statistical data modeling methods. Assessment of the efficiency of selected solutions.

Advanced data compression. Assessment of the efficiency and computational complexity of the indicated techniques.
Statistical analysis of data. Methods of classification and grouping of data. Assessment of the efficiency of selected methods.
Data modeling using regression. Assessment of the efficiency of the developed model.
Data processing using machine learning methods and artificial neural networks.

Teaching methods

1. Lecture
Classes with distinct elements of a traditional lecture, a problem-based lecture (discussion of a specific problem with students) and a conversational lecture (mobilizing students to discuss a specific topic), depending on the content of the material presented. Selected lecture contents are presented on a multimedia projector or blackboard. The discussion of issues is accompanied by information about their practical application.
2. Laboratories
Computer classes using software that enables advanced analysis, processing and compression of data. Solving problems given by the instructor and/or defined in the laboratory manual. Interpretation of the obtained solution and formulation of conclusions. Discussion of the possibilities of practical application of the issues covered by the laboratory.

Bibliography

Basic:
1. Damian Karwowski, Zrozumieć Kompresję Obrazu, ISBN: 978-83-953420-0-4, Poznań 2019, Wydanie pierwsze (pełna wersja książki dostępna jest na stronie internetowej: www.zrozumieckompresje.pl).
2. M. Domański, Obraz cyfrowy, WKŁ, 2011.
3. K. Sayood, Kompresja danych - wprowadzenie, Wydawnictwo RM, 2002.
4. Tadeusz Morzy, Eksploracja danych, Wydawnictwo Naukowe PWN, 2023.
5. Jacek Tabor Marek Śmieja Łukasz Struski Przemysław Spurek Maciej Wołczyk, Głębokie uczenie. Wprowadzenie, Wydawnictwo Helion, 2022.

Additional:
1. David Salomon, Giovanni Motta, Handbook of data compression, Springer-Verlag London Limited 2010, ISBN 978-1-84882-902-2.
2. Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep learning, 2016 Massachusetts Institute of Technology, ISBN 9780262035613.

Breakdown of average student's workload

	Hours	ECTS
Total workload	103	4,00
Classes requiring direct contact with the teacher	38	1,50
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	65	2,50