University of Gdańsk

Subject card

Subject name and code	Techniques of Multidimensional Data Mining, PG_00193526		
Field of study	Bioinformatics		
Date of commencement of studies	October 2026	Academic year of realisation of subject	2027/2028
Education level	Bachelor's studies	Subject group	Obligatory subject group in the field of study Subject group related to scientific research in the field of study
Mode of study	full-time studies	Mode of delivery	at the university
Year of study	2	Language of instruction	Polish
Semester of study	4	ECTS credits	4.0
Learning profile	academic	Assessment form	exam
Conducting unit	Faculty of Chemistry -> Rector		
Name and surname of lecturer (lecturers)	Subject supervisor	prof. dr hab. Tomasz Puzyn	
	Teachers		

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	15.0	0.0	45.0	0.0	0.0	60

E-learning hours included: 0.0

Learning activity and number of study hours

Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
Number of study hours	60	0.0	40.0	100

Subject objectives

To introduce students to the range of applicability of multivariate data mining techniques in bioinformatics, in particular in the analysis of genomic, trascriptomic, proteomic and metabolomic and chemomic data.To provide students with the skills to use the most important multivariate data mining techniques (correct method selection, performing analyses and interpreting the obtained results).To familiarise students with the capabilities of Python libraries for multivariate data mining techniques.Translated with DeepL.com (free version)

Learning outcomes

Course outcome	Subject outcome	Method of verification
[BIOINL3_U03] Graduate applies mathematical and statistical methods to describe phenomena and analyze data; has the ability to perform data analysis in professional databases used in bioinformatics	The student is able to correctly formulate a research problem (research question) and select an appropriate multidimensional data mining technique for it. The student is able to correctly carry out an analysis of the internal structure of a set of multidimensional data on the basis of data mining techniques with the use of tools available for the Python language and independently programmed scripts, and to correctly interpret the obtained results. The student is able to correctly present (in the form of a written report) the conducted analyses and discuss the obtained results.	[SU2] presentation/project/paper/ report [SU3] text preparation/written work [SU4] test/exam - oral or written
[BIOINL3_W03] Has sufficient knowledge of mathematical and statistical methods in order to describe and model biological phenomena and processes	The student will know the theoretical basis (algorithm of operation) of the most important multidimensional data mining techniques. The student will identify examples of the application of multidimensional data mining techniques in bioinformatics.	[SW4] test/exam - oral or written [SW1] oral statement/ conversation/discussion [SW2] presentation/project/paper/ report

Subject contents

Specificities of big data; Archiving of multidimensional data; Data repositories; Data import/export; Pre-processing and pre-screening methods; Visualisation of Big data; Archiving of multidimensional data; Data repositories; Data import/export; Methods of data pre-checking and pre-processing; Visualisation of multidimensional data; Similarity of objects in a multidimensional feature space; Methods for expressing similarity between objects and between objects and clusters of objects (Euclidean distance, Chebyshev, urban, complementary, single and full bound methods); The problem of dimensionality reduction of a dataset; Redundancy of information; Divergent objects in high-dimensional space; Unsupervised machine learning algorithms: One-way and two-way hierarchical cluster analysis (HCA), Principal component analysis (PCA), Methods for selecting variables and representative objects.In this class, special emphasis will be placed on the practical application of multivariate data mining techniques in the analysis of genomic, transcriptomic, metabolomic and chemomic data (comparison of structural features and properties of chemical compounds in large sets). Modelling of transcriptomic and metabolic pathways. Identification of disease markers at the molecular level based on multidimensional proteomic and metabolomic data. Application of multidimensional data mining techniques in personalised medicine. Scripting in Python. Combining available tools (R, Python, KNIME).Translated with DeepL.com (free version)

Prerequisites and co-requisites

Mathematical methods for bioinformaticsComputer science - introductionNumerical methods for bioinformaticsPython with fundamentals of algorithmicsStatistical analysis and probability calculus for bioinformaticians

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
exam	51.0%	100.0%

Recommended reading

Basic literature	• J. Mazerski: Podstawy chemometrii. Gdańsk: Wydawnictwo Politechniki Gdańskiej, 2000 • M. Gągolewski: Programowanie w języku R. PWN, 2016 • M. Lutz: Python. Wprowadzenie. Helion, 2002 • S. Raschka: Python. Uczenie maszynowe. Helion, 2016
Supplementary literature	Scientific publications
eResources addresses	

Example issues/ example questions/ tasks being completed

Work placement

Not applicable

Document generated electronically. Does not require a seal or signature.