



Subject card

Subject name and code	Data mininig methods, PG_00064605						
Field of study	Metody eksploracji danych						
Date of commencement of studies	February 2026		Academic year of realisation of subject		2025/2026		
Education level	second-cycle 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	1		Language of instruction		Polish		
Semester of study	1		ECTS credits		4.0		
Learning profile	general academic profile		Assessment form		assessment		
Conducting unit	Department of Theoretical Physics and Quantum Computing -> Faculty of Applied Physics and Mathematics -> Faculties of Gdańsk University of Technology						
Name and surname of lecturer (lecturers)	Subject supervisor		dr inż. Michał Piłat				
	Teachers						
Lesson types	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	0.0	0.0	30.0	45
	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	45		5.0		50.0	100
Subject objectives	Teach students basic and advanced methods for the analysis of large data sets. Teach a critical approach to the results you receive. Teach the preparation of presentations containing theoretical content and relevant examples.						
Learning outcomes	Course outcome		Subject outcome		Method of verification		
	[K7_W04] has enhanced knowledge of mathematical, numerical and simulation methods applied in the description and modelling of physical phenomena		Has in-depth knowledge of mathematical, numerical and simulation methods used in the description and modeling of data.		[SW2] Ocena wiedzy zawartej w prezentacji		
	[K7_U06] can apply obtained knowledge of physics to exact sciences, natural and technical sciences		Can apply the acquired knowledge in the field of physics to issues in the area of other sciences, natural sciences or technical sciences.		[SU1] Ocena realizacji zadania		
	[K7_U07] has enhanced skill of preparing speeches in Polish and English, including presentation of own research results		Can deliver an oral presentation on a chosen topic.		[SU5] Ocena umiejętności zaprezentowania wyników realizacji zadania		

Subject contents	Course content – lecture Lecture: 1. Definitions, stages and aims of data exploration process. Basic structures and models used for data exploration. Using algorithms in exploration of big data. • 2. Basic definitions of descriptive statistics, including tests and confidence intervals 3. Data preprocessing. Data Cleaning and handling missing data. Removing variables that are not useful. Choosing independent variables. Exploratory Data Analysis. • 4. Modelling of linear regression. The least squares method. Correlation and determination coefficients. 5. Modelling of multiple regression. Model assumptions and verifications. • 6. Logistic regression. Estimation of highest reliability. Interpretation of results. Conclusions on reliability of independent variables. 7. Naive Bayes and Bayesian networks. The Maximum Posteriori classification. The posteriori odds ratio. Balancing the data. Naive Bayes classification. • 8. Supervised and unsupervised methods. Methodology of supervised modelling. k-nearest neighbour algorithm 9. Decision trees. Application of the C4.5 i CART algorithms to real data. Decision rules. • 10. Neural networks. Encoding of input and output data. Sigmoid activation function. Learning rate. 1. 11. The Kohonen networks. 1. <		
------------------	---	--	--

Recommended reading	Basic literature	Literature: Daniel T. Larose, Discovering Knowledge in Data. An Introduction to Data Mining, John Wiley & Sons, Inc Daniel T. Larose, Data Mining Methods and Models, John Wiley & Sons, Inc
	Supplementary literature	Internet pages
	eResources addresses	
Example issues/ example questions/ tasks being completed	1 . Describe the similarities and differences between neural networks, Kohonen networks and limited Boltzmann machines. 2. What advice would you give to a person who is proceeding to data mining? 3. Build a decision tree using any algorithm that determines the credit risk based on the given data:	
Practical activities within the subject	Not applicable	

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