



Subject card

Subject name and code	Data Warehousing, PG_00044140						
Field of study	Mathematics						
Date of commencement of studies	October 2022		Academic year of realisation of subject		2023/2024		
Education level	second-cycle studies		Subject group		Optional subject group 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	3		ECTS credits		4.0		
Learning profile	general academic profile		Assessment form		assessment		
Conducting unit	Faculty of Applied Physics and Mathematics						
Name and surname of lecturer (lecturers)	Subject supervisor		dr hab. Paweł Pilarczyk				
	Teachers		dr hab. Paweł Pilarczyk mgr inż. Michał Krzemiński				
Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	30.0	0.0	30.0	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		5.0		35.0	100
Subject objectives	Theoretical and practical introduction to data mining and data warehousing.						
Learning outcomes	Course outcome		Subject outcome		Method of verification		
	K7_U13		The student knows mathematical foundations of the data mining algorithms that he or she is familiar with, including machine learning, and is capable of applying these algorithms to specific problems.		[SU4] Assessment of ability to use methods and tools [SU1] Assessment of task fulfilment		
	K7_U08		The student can conduct basic statistical analysis of numerical quantitative and qualitative data.		[SU3] Assessment of ability to use knowledge gained from the subject [SU2] Assessment of ability to analyse information		
	K7_W10		The student can choose statistical methods and machine learning algorithms, as well as software environment appropriate for modeling real world problems.		[SW3] Assessment of knowledge contained in written work and projects		
	K7_K02		The student can plan a data mining project, and then discuss its outcome.		[SK4] Assessment of communication skills, including language correctness		
	K7_W08		The student knows mathematical methods and software tools used in data mining.		[SW1] Assessment of factual knowledge		
Subject contents	Lecture: introduction to data mining and knowledge discovery in data. Data preprocessing and exploratory data analysis, Cross-Industry Standard Process for Data Mining (CRISP-DM). Statistical data analysis and machine learning. Methods for classification and data clustering, discovering association rules. Data warehousing, multidimensional modeling, OLAP. Laboratory: practical data mining and data exploration using Python and R programming.						

Prerequisites and co-requisites	Basic ability to write programs in R and in Python.		
	Familiarity with basic statistical methods.		
Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	Group projects (laboratory)	60.0%	50.0%
	Quizzes (lecture)	60.0%	50.0%
Recommended reading	Basic literature	Daniel T. Larose, Chantal D. Larose. Discovering Knowledge in Data. An Introduction to Data Mining, 2nd Ed., 2014.	
		Daniel T. Larose. Data Mining Methods and Models. Wiley-IEEE Press; 1st Ed., 2006.	
		Jiawei Han, Micheline Kamber, Jian Pei. Data Mining. Concepts and Techniques. 3rd Ed. Elsevier, 2011.	
	Supplementary literature	Wes McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media; 2nd Ed., 2017.	
		Joel Grus. Data Science from Scratch: First Principles with Python. O'Reilly Media; 1st Ed., 2015.	
		Jacek Rumiński. Wprowadzenie do hurtowni i eksploracji danych. Gdańsk, Wydawnictwo Politechniki Gdańskiej, 2015.	
	eResources addresses	Adresy na platformie eNauczanie: Hurtownie danych 2023 - Moodle ID: 29955 https://enauczanie.pg.edu.pl/moodle/course/view.php?id=29955	
Example issues/ example questions/ tasks being completed	Methods for data preprocessing.		
	What is the difference between a data warehouse and an operational database?		
	Methods for supervised data classification.		
	Kohonen networks and their relation to neural networks.		
Work placement	Not applicable		