



Subject card

Subject name and code	Data collection and analysis, PG_00045766						
Field of study	Technical Physics						
Date of commencement of studies	February 2023	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	1	Language of instruction			Polish		
Semester of study	2	ECTS credits			3.0		
Learning profile	general academic profile	Assessment form			assessment		
Conducting unit	Department of Theoretical Physics and Quantum Information -> Faculty of Applied Physics and Mathematics						
Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Bartosz Reichel					
	Teachers	dr inż. Bartosz Reichel					
Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	30.0	0.0	0.0	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		25.0	75
Subject objectives	<ol style="list-style-type: none">1. Practical introduce to capture video stream2. Test the capabilities and actions of mapped files in comparison to the normal files3. Familiar with serial communication on the example of UART with software and hardware flow control4. Introduction to the real aspects of data analysis in business applications5. Introduction to the basic notions of data analysis and data mining6. Mastering the skills of selection of tools and techniques needed to analyze and explore data						
Learning outcomes	Course outcome	Subject outcome			Method of verification		
	[K7_K05] Can communicate and present results of own work and transfer information in a commonly understandable manner.	Student is aware of and understands the validity of non-technical aspects and effects of engineering activities, including its impact on the environment, and the associated responsibility for decisions.			[SK5] Assessment of ability to solve problems that arise in practice		
	[K7_U02] Has enhanced knowledge of programming languages and can use software packages.	Uses packets during laboratory.			[SU1] Assessment of task fulfilment		
	[K7_W04] Has enhanced knowledge of mathematical, numerical and simulation methods applied in the description and modelling of physical phenomena.	Uses during the implementation of tasks (e.g. network traffic analysis, data modeling)			[SW1] Assessment of factual knowledge		

Subject contents	<p>The main goal of the course is to introduce students to fundamentals of data mining and data analysis. The course will provide examples of real-world problems in the fields of business, medicine and IT.</p> <p>They will be presented to the processes of collecting and recording data encountered in practice-row data collection video, audio, data transferred via serial interfaces, data acquisition cards.</p> <ol style="list-style-type: none"> 1. Video data collection 2. Serial interfaces: UART, USB, Ethernet 3. Mapped files, blocking 4. Database SQL / NoSQL, storage optimization 5. Collecting data about web user behavior 6. Collecting data about the user's location <ol style="list-style-type: none"> 7. Streaming (eg Apache Kafka) <ol style="list-style-type: none"> 8. MQTT as a data collection control protocol for IoT, MQTT Brokers <p>Emphasis will be placed on how to solve real problems during analytical process. Scope of the course includes:</p> <ol style="list-style-type: none"> 1. Data analysis as a process. Data mining methodologies. 2. Data preparation. 3. Classification techniques: Naïve Bayes classifie, k-NN 4. Regression methods 5. Classification and regression trees. Optimalization. <ol style="list-style-type: none"> 6. Assessing and comparing models performance <ol style="list-style-type: none"> 7. Machine learning methods, types of layers in neural networks. 											
Prerequisites and co-requisites												
Assessment methods and criteria	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 33%;">Subject passing criteria</th> <th style="width: 33%;">Passing threshold</th> <th style="width: 33%;">Percentage of the final grade</th> </tr> </thead> <tbody> <tr> <td>Laboratory</td> <td>50.0%</td> <td>50.0%</td> </tr> <tr> <td>Egzam</td> <td>50.0%</td> <td>50.0%</td> </tr> </tbody> </table>			Subject passing criteria	Passing threshold	Percentage of the final grade	Laboratory	50.0%	50.0%	Egzam	50.0%	50.0%
Subject passing criteria	Passing threshold	Percentage of the final grade										
Laboratory	50.0%	50.0%										
Egzam	50.0%	50.0%										
Recommended reading	Basic literature	<ol style="list-style-type: none"> 1. D. Hand, H. Mannila, P. Smyth, Principles of data mining, MIT Press, 2001 2. F. Provost, T. Fawcett, Data science for business, O'Reilly Media, 2013 										
	Supplementary literature	J.Axelson, Serial Port Complete: COM Ports, USB Virtual COM Ports, and Ports for Embedded Systems (Complete Guides series), 2007, Lakeview Research										
	eResources addresses	Adresy na platformie eNauzanie:										
Example issues/ example questions/ tasks being completed	<p>Implementation of the circular buffer.</p> <p>Describe the process for data warehouse ETL?</p>											
Work placement	Not applicable											