



Subject card

Subject name and code	Data engineering tools, PG_00045768						
Field of study	Technical Physics						
Date of commencement of studies	February 2024	Academic year of realisation of subject			2024/2025		
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	Instytut Fizyki i Informatyki Stosowanej -> Faculty of Applied Physics and Mathematics						
Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Paweł Syty					
	Teachers	dr inż. Paweł Syty					
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	To familiarize students with selected IT systems for data processing and analysis.						
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.	The student has in-depth knowledge of data processing methods in selected information systems and is able to create a mathematical model of a selected fragment of reality using selected methods of data engineering and machine learning.			[SW1] Assessment of factual knowledge		
	[K7_U02] Has enhanced knowledge of programming languages and can use software packages.	The student is able to use selected IT systems together with the accompanying programming languages to analyze information and process it.			[SU2] Assessment of ability to analyse information [SU4] Assessment of ability to use methods and tools [SU1] Assessment of task fulfilment		
Subject contents	Review of available data engineering tools. Advantages and disadvantages of selected tools. Introduction to machine learning on the example of selected software. Apache Hadoop as a tool for distributed storage and processing of large data sets using computer clusters. Use cases using Apache Spark and Apache Hive. Examples of using the SAS environment in data engineering: SAS Data Integration Studio, SAS Enterprise Miner. Exemplary possibilities of using H2O software in Big Data. Introduction to advanced business analytics on the example of Power BI.						
Prerequisites and co-requisites	Knowledge of data mining methods. Knowledge of Python and R programming languages.						
Assessment methods and criteria	Subject passing criteria	Passing threshold			Percentage of the final grade		
	Project	50.0%			100.0%		
Recommended reading	Basic literature	Trevor Hastie, Robert Tibshirani, Jerome Friedman, „The Elements of Statistical Learning: Data Mining, Inference, and Prediction”, Springer 2008. Alan Agresti, “An Introduction to Categorical Data Analysis”, Wiley - Interscience 2007.					
	Supplementary literature	“SAS Visual Analytics 7.4: User’s Guide”, SAS Institute Inc. 2015					
	eResources addresses	Adresy na platformie eNauczanie:					

Example issues/ example questions/ tasks being completed	Analyze the selected data set using SAS Enterprise Miner.
Work placement	Not applicable