



Subject card

Subject name and code	Data engineering tools, PG_00045768						
Field of study	Technical Physics						
Date of commencement of studies	February 2027	Academic year of realisation of subject			2027/2028		
Education level	second-cycle studies	Subject group			Specialty 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	Institute of Physics and Applied Computer Science -> Faculty of Applied Physics and Mathematics -> Faculties of Gdańsk University of Technology						
Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Patryk Jasik					
	Teachers	dr inż. Patryk Jasik					
Lesson types	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 used for data processing and analysis, particularly large volumes of data. To acquire knowledge and skills related to the practical use of data engineering tools.						
Learning outcomes	Course outcome	Subject outcome			Method of verification		
	[K7_W04] possesses advanced knowledge of mathematical, numerical and simulation methods used in the description and modelling of physical phenomena.						
	[K7_W05] understands the physical principles underlying the operation of scientific equipment and the processes occurring throughout its life cycle.						
	[K7_U02] demonstrates advanced programming skills in a selected language and the ability to use specialised software packages.						
	[K7_U04] is able to formulate and test research hypotheses.						

Subject contents	Course content – lecture		
	<p>Overview of available data engineering tools. Advantages and limitations of selected data engineering tools. Apache Hadoop as a tool for distributed storage and processing of large datasets using computer clusters. The theoretical foundations of Apache Spark. Fundamentals of Physics-Informed Neural Networks (PINNs).</p>		
	Course content – laboratory		
	<p>Training machine learning models with Apache Spark. Solving selected physics problems using Physics-Informed Neural Networks (PINNs). Example applications of the H2O platform in big data. Introduction to MLflow. Introduction to Apache Airflow. Introduction to AutoGluon.</p>		
Prerequisites and co-requisites	Knowledge of data mining methods. Ability to program in Python.		
Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	Completion of selected specialized courses on the DataCamp portal	60.0%	70.0%
	Attendance at classes	80.0%	30.0%
Recommended reading	Basic literature	Ofer Mendelevitch, Casey Stella, Douglas Eadline, Practical Data Science with Hadoop and Spark, Addison-Wesley Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer 2008.	
	Supplementary literature	Alan Agresti, An Introduction to Categorical Data Analysis, Wiley - Interscience 2007.	
	eResources addresses		
Example issues/ example questions/ tasks being completed	DataCamp - Introduction to PySpark (4h) DataCamp - Introduction to MLflow (4h) DataCamp - Introduction to Apache Airflow in Python (4h)		
Practical activities within the subject	Not applicable		

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