GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Data engineering tools, PG_00045768		
Field of study	Technical Physics		
Date of commencement of studies	February 2026	Academic year of realisation of subject	2026/2027
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] has enhanced knowledge of mathematical, numerical and simulation methods applied in the description and modelling of physical phenomena	The student is able to use selected data engineering tools to describe and model physical phenomena, drawing on advanced knowledge of mathematical, numerical, and simulation methods.	[SW3] Assessment of knowledge contained in written work and projects
[K7_U04] can formulate and test hypotheses related to research problems	The student is able to formulate hypotheses related to research problems and verify them using selected data engineering tools.	[SU4] Assessment of ability to use methods and tools
[K7_K03] can communicate and present results of own work and transfer information in a commonly understandable manner	The student is able to present the results of his/her work obtained using selected data engineering tools.	[SK5] Assessment of ability to solve problems that arise in practice
[K7_U02] has enhanced knowledge of programming languages and can use software packages	The student has in-depth programming skills in Python using its packages within an ecosystem created using selected data engineering tools.	[SU4] Assessment of ability to use methods and tools

<table>
<tr><td>Subject contents</td><td>Course content – lecture

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).

Course content – laboratory

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.</td></tr>
<tr><td>Prerequisites and co-requisites</td><td>Knowledge of data mining methods. Ability to program in Python.</td></tr>
<tr><td>Assessment methods and criteria</td><td><table>
<tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr>
<tr><td>Attendance at classes</td><td>80.0%</td><td>30.0%</td></tr>
<tr><td>Completion of selected specialized courses on the DataCamp portal</td><td>60.0%</td><td>70.0%</td></tr>
</table></td></tr>
<tr><td>Recommended reading</td><td><table>
<tr><td>Basic literature</td><td>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.</td></tr>
<tr><td>Supplementary literature</td><td>Alan Agresti, An Introduction to Categorical Data Analysis, Wiley - Interscience 2007.</td></tr>
<tr><td>eResources addresses</td><td></td></tr>
</table></td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td>DataCamp - Introduction to PySpark (4h)
DataCamp - Introduction to MLflow (4h)
DataCamp - Introduction to Apache Airflow in Python (4h)</td></tr>
<tr><td>Practical activites within the subject</td><td>Not applicable</td></tr>
</table>

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