GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Machine learning, PG_00061911
Field of study	Materials Engineering

Date of commencement of studies	October 2023	Academic year of realisation of subject	2024/2025
Education level	first-cycle studies	Subject group	Obligatory subject group 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	4	ECTS credits	2.0
Learning profile	general academic profile	Assessment form	assessment

Conducting unit	Division of Theoretical Physics and Quantum Informaton -> Institute of Physics and Applied Computer Science -> Faculty of Applied Physics and Mathematics

Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Patryk Jasik
	Teachers	dr inż. Patryk Jasik

Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	0.0	0.0	30.0	0.0	0.0	30
	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	30	5.0	15.0	50

Subject objectives	The aim of the course is to teach students the practical use of Python in the field of machine learning.

Learning outcomes	Course outcome	Subject outcome	Method of verification
	[K6_W01] Has knowledge of selected branches of mathematics, useful for formulating and solving problems and describing mechanical and physical phenomena, and chemical processes.	The student knows how to write Python code using knowledge from selected areas of mathematics necessary for formulating and solving problems, as well as describing mechanical, physical, and chemical phenomena.	[SW3] Assessment of knowledge contained in written work and projects
	[K6_U01] Can properly use selected analytical, simulation and experimental methods, as well as devices for measuring the fundamental properties of materials and technological processes.	The student is able to use appropriately selected analytical, simulation, and machine learning methods in Python.	[SU1] Assessment of task fulfilment
	[K6_U03] Can critically analyze and evaluate the functioning – particularly in the context of materials engineering –existing technical solutions, particularly equipment, objects, systems, processes.	Using Python and machine learning methods, the student is able to perform a critical analysis of how these solutions operate and evaluate existing technical solutions.	[SU1] Assessment of task fulfilment

Subject contents	
	1. Basic types of data and operations on them. Conditions. Different types of loops. Exceptions. Lists, tuples, dictionaries. 2. Functions. Function with an optional argument. Modules. Creating own module and using it. Writing/ reading data to/from files. YAML files. Classes and objects. Introduction to multithreading. 3. Analysis of the selected dataset. Reading observations for selected variables. Checking basic statistics for individual variables. Creating histograms. Identification of variables with potentially incorrect data (observations) or missing data. Data repair. Calculation of normalized correlations between individual variables. Conducting linear regression for selected variables, including graphs. 4. The scikit-learn package and the linear regression model. The coefficient of determination R^2, MSE, MAE. Division of the data set into a training and test part. Prediction using the created model. 5. The scikit-learn and preprocessing package. Polynomial model. Feature engineering. Reduction of variables in the model - the Schwarz criterion (Bayesian Information Criterion - BIC). Operation of the polynomial model in practice. 6. The scikit-learn package, the k-nearest neighbors method, decision trees and random forests. Classification problem. The choice of features - predictors and the target variable. Parameters of the model. Model quality assessment - confusion matrix, sensitivity, specificity, precision, accuracy, ROC curve, LIFT curve. Cross-validation: k-fold, n-fold and Monte-Carlo (bootstrap). 7. The scikit-learn package and the k-means algorithm as a case of unsupervised learning. Cluster analysis - clustering. Parameters of the model. The Fowlkes-Mallows index, i.e. the consistency between the two divisions of the dataset into clusters. Analysis of the principal components - PCA. 8. XGBoost, LightGBM packages as examples of reinforcement learning. 9. Introduction to neural networks using TensorFlow and PyTorch.

Prerequisites and co-requisites	Basic programming skills in the selected language. Basic knowledge of probability calculus and statistics.

Assessment methods and criteria			
	Subject passing criteria	Passing threshold	Percentage of the final grade
	The report of the analysis and modeling of the selected dataset	60.0%	100.0%

Recommended reading		
	Basic literature	1. Andreas C. Maźller, Sarah Guido, "Introduction to Machine Learning with Python. A Guide for Data Scientists", O'Reilly Media 2. Alberto Boschetti, Luca Massaro, "Python Data Science Essentials - Second Edition", Packt Publishing 3. Sebastian Raschka, "Python Machine Learning", Packt Publishing
	Supplementary literature	1. James R. Parker, "Python. An Introduction to Python Programming", Mercury Learning and Information 2. Krishna Kumar Mohbey, Brijesh Bakariya, "An Introduction to Python Programming: A Practical Approach", BPB Publications 3. Wes McKinney, "Python for Data Analysis. 3rd Edition", O'Reilly Media
	eResources addresses	Adresy na platformie eNauczanie: Uczenie maszynowe - Moodle ID: 45055 https://enauczanie.pg.edu.pl/moodle/course/view.php?id=45055

Example issues/ example questions/ tasks being completed	Guidelines for creating the report: 1. Report Title 2. Introduction motivation, objectives 3. Data Description dataset structure, variables, origin 4. Description of the Data Preparation Process step-by-step approach 5. Data Analysis assumptions, brief overview of methods and chosen analysis methodology 6. Data Modeling assumptions, brief overview of methods and chosen modeling methodology 7. Results, Conclusions, and Discussion The report, along with all code, should be placed in a chosen repository (e.g., GitLab, GitHub).

Work placement	Not applicable

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