GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Advanced Data Preparation in Machine Learning, PG_00054186		
Field of study	Informatics		
Date of commencement of studies	February 2024	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	1	ECTS credits	2.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Computer Architecture -> Faculty of Electronics, Telecommunications and Informatics		
Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Jan Cychnerski	
	Teachers	dr inż. Jan Cychnerski	

Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	15.0	0.0	0.0	15.0	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	2.0	18.0	50

Subject objectives	The main goal of the subject is to present practical methods of solving problems using artificial intelligence techniques: dataset construction, architecture choice, artificial intelligence algorithms training, model selection and testing.

Learning outcomes	Course outcome	Subject outcome	Method of verification
	[K7_W04] Knows and understands, to an advanced extent, the principles, methods and techniques of programming and the principles of computer software development or programming devices or controllers using microprocessors or programmable elements or systems specific to the field of study, and organisation of systems using computers or such devices	The student has the advanced knowledge necessary to design systems based on artificial intelligence. The student understands the principles of operation and ways of using commonly used libraries and environments providing self-learning algorithms. The student knows the practical limitations and the best ways to apply methods and systems of artificial intelligence. The student understands the reasons of these limitations and their practical effects.	[SW1] Assessment of factual knowledge
	[K7_U05] can plan and conduct experiments related to the field of study, including computer simulations and measurements; interpret obtained results and draw conclusions	The student is able to conduct training and performance tests of a neural network or other artificial intelligence algorithm.	[SU1] Assessment of task fulfilment
	[K7_W42] Knows and understands, to an increased extent, the principles and trends in the analysis and design of local and distributed IT systems and the basics of computer modeling and computerization of complex cognitive and decision-making processes.	The student has in-depth knowledge of mechanics of artificial intelligence systems. He understands the training and testing methods of self-learning algorithms. Student knows how train and test data sets should be constructed in order to achieve optimal results. Student understands how tests should be carried out to minimize errors. The student has the knowledge necessary to effectively and efficiently conduct the training process of commonly used classifiers based on artificial intelligence methods	[SW1] Assessment of factual knowledge
	[K7_U41] can select methods of modelling and analysis of information systems and applications using selected elements of theoretical computer science and modern programming tools	The student is able to select the methods of machine learning in the context of the requirements of the problem being solved	[SU1] Assessment of task fulfilment
	[K7_K02] is ready to provide critical evaluation of received content and to acknowledge the importance of knowledge in solving cognitive and practical problems	The student is able to assess the difficulty of problems solved by machine learning methods and search for information helpful in solving them.	[SK5] Assessment of ability to solve problems that arise in practice

Subject contents	1. General information about practical aspects of training artificial intelligence algorithms 2. Environments, tools and helper libraries in machine learning 3. Preparation of training, validation and test data sets 4. Preprocessing, normalization and augmentation of training data 5. Choosing machine learning methods in the context of problem requirements 6. Methods of performing training of artificial intelligence algorithms 7. Methods of machine learning hyperparameter assignment 8. Methods of testing and measuring effectiveness and performance of artificial intelligence algorithms 9. Identifying and solving typical problems in machine learning 10. Deployment of machine learning algorithms in the target environment
Prerequisites and co-requisites	Basic knowledge of artificial intelligence area, basic knowledge of Python programming language

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
laboratory	50.0%	50.0%
written test	50.0%	50.0%

Recommended reading

Basic literature	1. James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013. 2. Ian Goodfellow, Yoshua Bengio and Aaron Courville, “Deep Learning”, http://www.deeplearningbook.org/ 3. Scikit-learn Tutorials, http://scikit-learn.org/stable/tutorial/index.html
Supplementary literature	1. Andrew Ng, “Machine Learning Yearning”, http:// www.mlyearning.org/
eResources addresses	Adresy na platformie eNauczanie: Zaawansowane przygotowanie danych w uczeniu maszynowym - 2023/2024 - Moodle ID: 36791 https://enauczanie.pg.edu.pl/moodle/course/view.php?id=36791

Example issues/ example questions/ tasks being completed	Sample issues: - Training and testing datasets preparation for classifier training - Performing training and testing of a neural network or other artificial intelligence algorithm - Description of methods and measures for quality, efficiency and performance of machine learning algorithms
Work placement	Not applicable