University of Gdańsk

Subject card

Subject name and code	Introduction to Machine Learning, PG_00143545		
Field of study	Informatics		
Date of commencement of studies	October 2025	Academic year of realisation of subject	2026/2027
Education level	Bachelor's studies	Subject group	Obligatory subject group in the field of study 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	2	Language of instruction	Polish
Semester of study	4	ECTS credits	4.0
Learning profile	academic	Assessment form	credit
Conducting unit	Division of Quantum Computing -> Institute of Informatics -> Faculty of Mathematics, Physics and Informatics -> Rector		
Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Paweł Mazurek	
	Teachers		

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	30.0	0.0	30.0	0.0	0.0	60

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	60	0.0	40.0	100

Subject objectives

1. Obtaining the basic methods of machine learning

2. Acquiring the ability to use basic machine learning algorithms - working in the Python environment

3. Mastering mathematical methods of modeling problems and machine learning tools and assessing their usefulness

Learning outcomes

Course outcome	Subject outcome	Method of verification
[INFL3_W03] has structured, theoretically grounded general knowledge in the field of algorithms and data structures, formal languages, automata theory and computational complexity, and artificial intelligence	1. The student is able to use the Python environment for machine learning. 2. The student knows data analysis methods. 3. The student is able to use the TensorFlow library tools to train neural networks.	[SW4] test/exam - oral or written [SW2] presentation/project/paper/ report
[INFL3_W10] knows the basic principles of occupational health and safety in the IT profession	Analysis of the workplace in the laboratory room in terms of occupational health and safety rules.	[SW4] test/exam - oral or written

Subject contents	1.Introduction to basic terminology and notation. System creation strategy in machine learning2. Python in machine learning. Using the Python environment for machine learning: Anaconda, Jupyter, NumPy, Pandas, Matplotlib, SciPy.3.Data pre-processing. Best model evaluation methods and parametric tuning.4. Data analysis using regression analysis, cluster analysis5.Training neural networks using the TensorFlow library6. Application examples:(i) Modeling sequential data using recurrent neural networks(ii) Working with text(iii) Working with images	
Prerequisites and co-requisites	Knowledge of the basics of programming, algorithms and complexity, languages and paradigms of object-oriented programming in the languages chosen by the student.	
Assessment methods and criteria		

Subject passing criteria	Passing threshold	Percentage of the final grade
a test exam of lecture content	51.0%	30.0%
graded project	51.0%	70.0%

Recommended reading	Basic literature	1. Python Machine Learning By Example, Yuxi Liu, Packt Publishing, 2024 2. M. Szeliga, Data science i uczenie maszynowe, Wydawnictwo naukowe PWN, Warszawa, 2017. 3. T. Morzy, Eksploracja danych metody i algorytmy, Wydawnictwo naukowe PWN, Warszawa, 2013. 4. Christopher M. Bishop Pattern Recognition and Machine Learning, Springer, 2007
	Supplementary literature	1. J. Koronacki, J. Ćwik: Statystyczne systemy uczące się. Wydanie drugie, EXIT, Warszawa, 20072. K. Krawiec, J.Stefanowski, Uczenie maszynowe i sieci neuronowe. Wydawnictwo Politechniki Poznańskiej, Poznań, 2004,3. P. Cichosz, Systemy uczace sie, WNT, Warszawa, 2000.4. W. Duch, J. Korbicz, L.Rutkowski,R. Tadeusiewicz, Sieci neuronowe, Exit, Warszawa, 20005. M. Gągolewski, M. Bartoszuk, A. Cena, Przetwarzanie i analiza danych w języku Python, PWN, 20166. W. McKinney, Python for Data Analysis. Data Wrangling with Pandas, NumPy, and IPython, O'Reilly Media, 20127. W. Richert, L.P. Coelho, Building Machine Learning Systems with Python, Packt Publishing, 20138. M. Lutz, Learning Python, O'Reilly Media, 20139.. E. Bressert, SciPy and NumPy, O'Reilly Media, 2012
	eResources addresses	

Example issues/ example questions/ tasks being completed	1. **Types of machine learning** (supervised, unsupervised, and reinforcement learning). 2. **Overfitting and underfitting**: definitions and detection methods. 3. **Cross-validation**. 4. **Regularization**. 5. **Feature selection and dimensionality reduction**. 6. **Feature engineering**, including one-hot encoding and dense embeddings. 7. **Prediction performance metrics** (precision, recall, AUC, etc.). 8. **Bayesian theory and its applications in machine learning**; best practices, including handling class imbalance, smoothing techniques, approaches for small and large datasets, and typical model parameters. 9. **Decision tree learning algorithms**, **random forest algorithms**, and **gradient-boosted tree algorithms**. Typical parameters of decision trees, random forests, and boosted trees. Advantages and disadvantages of decision trees, random forests, and boosted trees. 10. **Information measures** (Gini impurity, entropy, etc.). 11. **Logistic regression as an optimization problem**. Existence of analytical solutions depending on the type of regularization. Typical model parameters. 12. **Logistic regression for multiclass classification**. 13. **Linear regression and regression trees**. Differences compared with classification trees and logistic regression. 14. **Neural networks**: typical parameters, activation functions, methods for preventing overfitting (dropout, early stopping), and best practices for hyperparameter tuning. 15. **Tokenization, part-of-speech tagging, stemming, lemmatization, and stop words**. 16. **Visualization of high-dimensional data using t-SNE**. 17. **Dimensionality reduction using Principal Component Analysis (PCA)**. 18. **Bag-of-Words, TF-IDF, and dense embeddings**: advantages, disadvantages, and typical applications. 19. **The k-means clustering algorithm**: typical parameters and best practices. 20. **Advanced topic modeling methods** (Non-negative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA)). 21. **Support Vector Machines (SVMs) as a semidefinite optimization problem**. 22. **Support Vector Machines for multiclass classification**.

	23. **Linear and nonlinear kernels in Support Vector Machines**: applications, best practices, and typical parameters. **Convolutional Neural Networks (CNNs)**: architecture, typical parameters, comparison with fully connected neural networks, sources of nonlinearity and convolution, pooling functions, and best practices.
Work placement	Not applicable

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