GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	MACHINE LEARNING, PG_00070575		
Field of study	Economic Analytics		
Date of commencement of studies	October 2026	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	Part-time studies	Mode of delivery	at the university
Year of study	2	Language of instruction	Polish
Semester of study	3	ECTS credits	3.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Statistics and Econometrics -> Faculty of Management and Economics -> Faculties of Gdańsk University of Technology		
Name and surname of lecturer (lecturers)	Subject supervisor	dr hab. Michał Pietrzak	
	Teachers		

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	9.0	0.0	18.0	0.0	0.0	27

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	27	3.0	45.0	75

Subject objectives

to introduce students to the fundamental concepts, techniques, and algorithms used in machine learning for data analysis, prediction, and decision-making. Students will acquire both theoretical knowledge and practical skills in applying supervised and unsupervised learning methods, data preprocessing, model validation, and performance evaluation. The course emphasizes understanding the machine learning workflow, interpreting models, and applying them to real-world problems across various domains.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K7_U01] creates innovative solutions for complex and unstructured processes, considering unpredictable environmental conditions through the synthesis of information from various sources.	is able to design and implement machine learning models to analyze complex processes under changing environmental conditions. They can assess the relevance of various data sources and integrate them to obtain consistent and accurate conclusions.	[SU2] Assessment of ability to analyse information [SU4] Assessment of ability to use methods and tools [SU1] Assessment of task fulfilment
[K7_W03] demonstrates in-depth knowledge of the applications of analytical methods and techniques for formulating and solving socio-economic problems.	has advanced knowledge of data analysis methods and machine learning algorithms, enabling the identification and modeling of complex and unstructured processes. They are familiar with modern tools and technologies used for processing data from heterogeneous sources.	[SW3] Assessment of knowledge contained in written work and projects
[K7_K01] is ready to critically evaluate his/her knowledge in economic analytics and seek expert opinions when facing difficulties in solving a problem independently.	is ready to critically assess machine learning models and analyses, recognising the limitations of applied methods and seeking expert support in complex situation	[SK3] Assessment of ability to organize work [SK5] Assessment of ability to solve problems that arise in practice

Subject contents

Course content – lecture

1. Introduction to Machine Learning definitions, scope and significance of ML, taxonomy of learning paradigms (supervised, unsupervised, reinforcement learning), and overview of the Python ecosystem for MLMathematical
2. Foundations for Machine Learning elements of linear algebra, calculus, statistics, and optimization methods used in model training
3. Data Preparation data cleaning techniques, feature transformations, exploratory data analysis (EDA), and data visualization methods
4. Regression Models linear regression and its extensions, logistic regression interpretation of model parameters, and regularization techniques (Ridge, Lasso)
5. Classification Methods overview of fundamental classification algorithms and their applications
6. Model Validation evaluation techniques, data splitting strategies, cross-validation, and analysis of overfitting and underfitting
7. Unsupervised Learning clustering methods, dimensionality reduction, and anomaly detection
8. Model Tuning hyperparameter optimization techniques and the concept of ML pipelines
9. Deep Learning theoretical foundations and architecture of neural networks
10. Machine Learning Applications real-world use cases and stages of an ML project pipeline
11. Ethics in Artificial Intelligence bias, model interpretability, and responsibility in AI systems

Course content – laboratory

1. Computational environment setup installation and configuration of the Python environment, dependency management (venv/conda), and use of Jupyter Notebook
2. NumPy and pandas libraries operations on data structures, processing and manipulation of tabular data
3. Exploratory data analysis data loading, analysis of dataset structure, and identification of missing values and variable types
4. Data preprocessing data cleaning, missing value imputation, categorical variable encoding, and feature scaling
5. Anomaly detection methods for identifying outlier observations
6. Data visualization use of graphical tools to analyze distributions and relationships between variables
7. Linear regression model implementation, parameter estimation, and interpretation of results
8. Extended regression models polynomial regression and regularization methods (Ridge, Lasso)
9. Logistic regression development and application of a classification model for binary variables
10. KNN and Naive Bayes algorithms implementation and comparative analysis of classification methods
11. Decision trees and Random Forest tree-based models and ensemble methods
12. Support Vector Machines (SVM) classification using kernel functions and analysis of decision boundaries
13. Model validation data splitting, cross-validation, and evaluation of model stability
14. Overfitting phenomenon analysis of the impact of model complexity on generalization ability
15. K-means clustering implementation and analysis of data grouping
16. DBSCAN and hierarchical clustering methods density-based and agglomerative algorithms
17. Principal Component Analysis (PCA) dimensionality reduction and variance analysis
18. Hyperparameter tuning application of Grid Search and Random Search methods

Prerequisites and co-requisites

Descriptive statistics, mathematical statistics, econometrics, forecasting, fundamentals of programming in R/ Python.

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
Final project	60.0%	50.0%
Computer lab test	60.0%	50.0%

Recommended reading

Basic literature

1. Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, NY: Springer.
2. Murphy, K. P. (2012). *Machine learning: A probabilistic perspective* . Cambridge, MA: MIT Press.
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.
4. Géron, A. (2022). *Hands-on machine learning with scikit-learn, Keras, and TensorFlow* (3rd ed.). Sebastopol, CA: OReilly Media.
5. Chollet, F. (2021). *Deep learning with Python* (2nd ed.). Shelter Island, NY: Manning Publications.
6. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). New York, NY: Springer.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York, NY: Springer.
8. Burkov, A. (2019). *The hundred-page machine learning book*. Quebec City, Canada: Andriy Burkov.
9. Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge, UK: Cambridge University Press.
10. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). Cambridge, MA: MIT Press.

	Supplementary literature	1. Molnar, C. (2022). *Interpretable machine learning* (2nd ed.). N.p.: Lulu.com. 2. McKinney, W. (2022). *Python for data analysis* (3rd ed.). Sebastopol, CA: OReilly Media. 3. VanderPlas, J. (2016). *Python data science handbook*. Sebastopol, CA: OReilly Media. 4. Brownlee, J. (2020). *Machine learning mastery with Python*. Melbourne, Australia: Machine Learning Mastery. 5. Sarker, I. H. (2021). *Machine learning: Algorithms and applications* . Boca Raton, FL: CRC Press. 6. Provost, F., & Fawcett, T. (2013). *Data science for business*. Sebastopol, CA: OReilly Media. 7. Grus, J. (2019). *Data science from scratch* (2nd ed.). Sebastopol, CA: OReilly Media. 8. Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). *Dive into deep learning*. Cambridge, UK: Cambridge University Press.
	eResources addresses	
Example issues/ example questions/ tasks being completed	1. Configure a Python environment (venv/conda), install required libraries, and launch Jupyter Notebook to prepare a data analysis environment. 2. Load a real-world dataset (e.g., CSV), perform exploratory data analysis (EDA), and identify missing values and potential anomalies in the data. 3. Prepare the dataset for modeling by imputing missing values, encoding categorical variables, and scaling features to obtain a machine-learning-ready dataset. 4. Build a linear regression model to predict a continuous variable, evaluate it using metrics (MSE, R^2), and interpret the results. 5. Extend the regression model using polynomial features and apply regularization (Ridge or Lasso) to improve generalization performance. 6. Implement classification models (logistic regression, KNN, Naive Bayes) and compare their performance on the same dataset. 7. Train Decision Tree, Random Forest, and SVM models, and analyze the impact of their parameters on prediction performance. 8. Apply train/test split and cross-validation to evaluate model stability and identify overfitting. 9. Perform clustering using K-means and DBSCAN, and compare the resulting groupings and their interpretation. 10. Apply PCA for dimensionality reduction and evaluate its impact on clustering or classification performance.	
Practical activites within the subject	Not applicable	

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