University of Gdańsk

Subject card

| | | | |
|---|---|---|---|
| Subject name and code | Machine Learning, PG_00193533 | | |
| Field of study | Bioinformatics | | |
| Date of commencement of studies | October 2026 | Academic year of realisation of subject | 2028/2029 |
| Education level | Bachelor's studies | Subject group | Obligatory subject group in the field of study<br>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 | 3 | Language of instruction | Polish |
| Semester of study | 5 | ECTS credits | 4.0 |
| Learning profile | academic | Assessment form | exam |
| Conducting unit | Laboratory of Environmental Chemoinformatics -> Department of Environmental Chemistry and Radiochemistry -> Faculty of Chemistry -> Rector | | |
| Name and surname of lecturer (lecturers) | Subject supervisor | dr Artur Mirocki | |
| | Teachers | | |

| Lesson types | Lesson type | Lecture | Tutorial | Laboratory | Project | Seminar | SUM |
|---|---|---|---|---|---|---|---|
| | Number of study hours | 15.0 | 0.0 | 45.0 | 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. Introducing students to the broad applications of machine learning methods in bioinformatics.<br>2. Students will learn to use essential machine learning methods (including correct method selection, variable selection, model training and validation, and result interpretation).<br>3. Students will become familiar with available software and scripting languages that implement these methods, specifically R, Python, and KNIME. |

## Learning outcomes

| Course outcome | Subject outcome | Method of verification |
|---|---|---|
| [BIOINL3_W04] Has advanced knowledge of research techniques and tools used in bioinformatics | The student understands the theoretical basis and operating algorithms of the most essential machine learning methods.<br>The student provides examples of machine learning applications in bioinformatics, pharmacy, and toxicology. | [SW4] test/exam - oral or written<br>[SW1] oral statement/ conversation/discussion<br>[SW2] presentation/project/paper/ report<br>[SW3] text preparation/written work |
| [BIOINL3_W03] Has sufficient knowledge of mathematical and statistical methods in order to describe and model biological phenomena and processes | The student understands the theoretical basis and operating algorithms of the most essential machine learning methods.<br>The student provides examples of machine learning applications in bioinformatics, pharmacy, and toxicology. | [SW4] test/exam - oral or written<br>[SW1] oral statement/ conversation/discussion<br>[SW2] presentation/project/paper/ report |
| [BIOINL3_U03] Graduate applies mathematical and statistical methods to describe phenomena and analyze data; has the ability to perform data analysis in professional databases used in bioinformatics | The student can correctly formulate a research problem (research question) and select an appropriate theoretical model and machine learning method.<br>The student can accurately build a dependency model, validate it using tools available for the Python language and custom scripts, and correctly interpret the results obtained.<br>The student can effectively present and discuss the modeling results in a written report. | [SU2] presentation/project/paper/ report<br>[SU3] text preparation/written work<br>[SU4] test/exam - oral or written |
| [BIOINL3_U02] Graduate is able to apply knowledge of natural sciences and science to formulate, analyze and solve problems related to bioinformatics | The student can correctly formulate a research problem (research question) and select an appropriate theoretical model and machine learning method.<br>The student can accurately build a dependency model, validate it using tools available for the Python language and custom scripts, and correctly interpret the results obtained.<br>The student can effectively present and discuss the modeling results in a written report. | [SU2] presentation/project/paper/ report<br>[SU3] text preparation/written work<br>[SU4] test/exam - oral or written |

## Subject contents

The supervised learning algorithms for modeling the dependencies between the target prediction output and the input features (regression and classification), including linear and multivariate linear regression (LR and MLR), logistic regression, principal components regression (PCR), partial least squares (PLS) regression, linear discriminant analysis (LDA), non-linear k-nearest neighbors (kNN) classifier, naive Bayes classifier, decision trees (DT), random forests, classification and regression trees (CART), support vector machines (SVM), and artificial neural networks (ANN).Methods for selecting the optimal set of variables in the model (i.e., stepwise selection, the uninformative variable elimination algorithm (UVE), and selection using a genetic algorithm (GA)); validation of regression and classification models (i.e., cross-validation using the leave-one-out method and external validation). Estimation of the boundaries of the model's applicability domain.

During the classes, particular emphasis will be placed on the practical application of machine learning to develop computational models for analyzing genomic, transcriptomic, and metabolomic data, Adverse Outcome Pathways (AOP) models, and structure-biological activity models used in pharmacy and toxicology. Additionally, students will learn to apply machine learning techniques in personalized medicine, create scripts in Python, and combine available tools such as R, Python, and KNIME.

<table>
<tr><td>Prerequisites and co-requisites</td><td colspan="2">Prerequisites: matrix calculus, problem-solving skills, analysis of function variability, knowledge of the Linux environment, basics of programming in object-oriented languages, knowledge of unsupervised machine learning methodsFormal requirements: mathematics, introduction to computer science, programming, multidimensional data mining techniques</td></tr>
<tr><td>Assessment methods and criteria</td><td colspan="2"><table><tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr><tr><td>Test</td><td>51.0%</td><td>100.0%</td></tr></table></td></tr>
<tr><td rowspan="3">Recommended reading</td><td>Basic literature</td><td>A. Required literature to pass the course (for the final exam):<br><br>A.1. Literature used during classes:<br><br>• Script for laboratory exercises prepared by the employees of the Environmental Chemoinformatics Laboratory Team<br><br>A.2. Literature independently studied by the student<br><br>• J. Mazerski: Podstawy chemometrii. Gdańsk: Wydawnictwo Politechniki Gdańskiej, 2000<br>• M. Gągolewski: Programowanie w języku R. PWN, 2016<br>• M. Lutz: Python. Wprowadzenie. Helion, 2002<br>• S. Raschka: Python. Uczenie maszynowe. Helion, 2016<br>• https://www.youtube.com/user/KNIMETV<br>• P. Biecek: Przewodnik po pakiecie R. Wrocław: Oficyna Wydawnicza GiS, 2014<br>• P. Biecek: Analiza danych z programem R. Modele liniowe z efektami stałymi, losowymi i mieszanymi: Warszawa PWN, 2020.</td></tr>
<tr><td>Supplementary literature</td><td>Additional literature<br><br>• S. D. Brown, R. Tauler, B. Walczak (red): Comprehensive chemometrics: Chemical and biochemical data analysis. Amsterdam: Elsevier, 2009 R. Kramer: Chemometric techniques for quantitative analysis. New York: Marcel Dekker, Inc, 2005.</td></tr>
<tr><td>eResources addresses</td><td></td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td colspan="2"></td></tr>
<tr><td>Work placement</td><td colspan="2">Not applicable</td></tr>
</table>

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