KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

Projekt współfinansowany przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Course title	ECTS code
Monographic lecture - Machine learning algorithms for small datasets	13.3.1310

Name of unit administrating study

null

Studies

faculty	field of study		
Wydział Chemii	Chemia	type	drugiego stopnia
		form	stacjonarne
		specialty	Digital Chemistry
		specialization	wszystkie

Teaching staff

dr Agnieszka Gajewicz-Skrętna

Forms of classes, the realization and number of hours

Forms of classes

Lecture

The realization of activities

classroom instruction

Number of hours

Lecture: 30 hours

ECTS credits

3
Lecture – 30 h
Student's own work – 30 h
Tutorial classes: 15 h
TOTAL: 75 h – 3 ECTS

The academic cycle

2024/2025 winter semester

Type of course

obligatory

Language of instruction

English

Teaching methods

- discussion
- multimedia-based lecture

Form and method of assessment and basic criteria for eveluation or examination requirements

Final evaluation

Graded credit

Assessment methods

Lecture – written test in a form of a set of questions.

The basic criteria for evaluation

Assessment criteria in accordance with the University of Gdańsk Study Regulations

a score of 51% or more required to pass the test

Method of verifying required learning outcomes

Required courses and introductory requirements

A. Formal requirements
none

B. Prerequisites
basic knowledge in chemistry and physics

Aims of education

familiarizing the students with real-world machine learning problems where available datasets suffer from small size, noise, missing values, and inconsistency
presenting the advantages and disadvantages of various types of machine learning algorithms dealing with limited by the amount (and quality) of the data available

Course contents

This course is designed to provide students with foundational knowledge on machine learning algorithms dealing with limited by the amount (and quality) of the data available, including:

the effects of data quality on machine learning algorithms (with particular emphasis on small, incomplete, noisy, imbalance or affected by artifacts datasets),
introduction to the bias-variance trade-off in machine learning,
overview of various types of machine learning algorithms to deal with small size and low-quality data (including kernel regression methods, such as lasso, elastic net and bridge regression, and their adaptive extensions),
introduction to the basic concepts and strategies of read-across (including averaging approach, similarity-weighted activity of nearest neighbors, filtering approach, search expansion approach),
real-world examples that illustrate a successful application of the machine learning algorithms for imperfect datasets in the applied sciences (e.g., chemical and pharmaceutical industries, environmental risk assessment of chemical substances).

Bibliography of literature

Literature required to pass the course

S. D. Brown, R. Tauler, B. Walczak (ed): Comprehensive chemometrics: Chemical and biochemical data analysis. Amsterdam: Elsevier, 2009

Extracurricular readings

J. Leszczynski, A. Kaczmarek-Kedziera, T. Puzyn, M. G. Papadopulos, H. Reis, M. Shukla (ed): Handbook of Computational Chemistry (2nd Edition). Springer 2016. Volume 5: Chemoinformatics, Puzyn T (ed.).

The learning outcomes (for the field of study and specialization)

K_W05:
has extended knowledge in the field of the specialisation studied
K_W06:
applies mathematics to the extent necessary to understand, describe and model chemical processes of extended complexity
K_W08:
demonstrates in-depth knowledge of theoretical computational and IT methods used to solve problems in chemistry
K_W09:
classifies specialist IT tools used in statistical evaluation of experiment results
K_U02
critically assesses the results of conducted, performed observations and theoretical calculations and discusses errors
K_U03
finds necessary information in specialist literature, databases and other sources, lists basic scientific journals in chemistry
K_K02
works in a team taking on various roles in it
K_K06
raises her/his professional and personal competences by using information provided in various sources

Knowledge

At the completion of this course, the student is expected to be able to:
know how the small size and low quality of a dataset impact traditional machine learning algorithms and provide few ways to mitigate these issues,
understand the bias-variance trade-off,
know the preprocessing techniques and various types of machine learning algorithms to deal with small size and low-quality data, where missing values and noise data are included,
know and understand the theoretical background of read-across approach,
describe the most important challenges for the application of machine learning for small size and low-quality data.

Skills

At the completion of this course, the student is expected to be able to:
properly implement modeling strategies for handling limited by the amount (and quality) of the data available,
choose and apply the appropriate machine learning algorithm to solve a particular problem under consideration in the chemistry science domain,
evaluate the efficacy of the developed model and critically interpret the results obtained with specific machine learning methods.

Social competence

At the completion of this course, the student is expected to be able to:
describe the benefits of using machine learning methods in her/his daily research practice,
understand the need of deeper learning of the machine learning,
develop interpersonal skills such as communication, cooperation in group (taking different roles), and problem-solving abilities,
understand the social aspects of practical use of knowledge and abilities as well as connected with them responsibility.

Contact

agnieszka.gajewicz@ug.edu.pl