**Nazwa przedmiotu:**

Introduction to Machine Learning

**Koordynator przedmiotu:**

Dr inż. Agnieszka Jastrzębska

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Data Science

**Grupa przedmiotów:**

Współny

**Kod przedmiotu:**

1120-INSZI-MSA-0667

**Semestr nominalny:**

1 / rok ak. 2022/2023

**Liczba punktów ECTS:**

4

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

1. godziny kontaktowe – 50 h; w tym
 a) obecność na wykładach – 15 h
 b) obecność na laboratoriach – 15 h
 c) obecność na zajęciach projektowych – 15 h
 d) konsultacje –5 h
2. praca własna studenta – 70 h; w tym
 a) zapoznanie się z literaturą – 10 h
 b) rozwiązanie zadań domowych (wykonanie projektu) – 30 h
 c) przygotowanie do zajęć laboratoryjnych – 10 h
 d) przygotowanie do zajęć projektowych – 10 h
 e) przygotowanie raportu/prezentacji – 10 h
Razem 120 h, co odpowiada 4 pkt. ECTS

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

1. obecność na wykładach – 15 h
2. obecność na laboratoriach – 15 h
3. obecność na zajęciach projektowych – 15 h
4. konsultacje – 5 h
Razem 50 h, co odpowiada 2 pkt. ECTS

**Język prowadzenia zajęć:**

angielski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

1. obecność na laboratoriach – 15 h
2. obecność na zajęciach projektowych – 15 h
3. rozwiązanie zadań domowych (wykonanie projektu) – 30 h
4. przygotowanie do zajęć laboratoryjnych – 10 h
5. przygotowanie do zajęć projektowych – 10 h
Razem 80 h, co odpowiada 3 pkt. ECTS

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład:  | 15h |
| Ćwiczenia:  | 0h |
| Laboratorium:  | 15h |
| Projekt:  | 15h |
| Lekcje komputerowe:  | 0h |

**Wymagania wstępne:**

Mathematics: algebra, calculus, probability theory, statistics, theoretical foundations of computer science: algorithms and data structures, programming

**Limit liczby studentów:**

Laboratory, project classes – the number of students in a group matches the limits defined by the Warsaw University of Technology

**Cel przedmiotu:**

The objective is to revise and synthesize fundamental information acquired from previous courses in mathematics and widely understood computational intelligence, to expand the scope of interest onto machine learning with a particular focus on practical abilities

**Treści kształcenia:**

Lecture:
Lectures cover elementary notions and techniques of the machine learning area:
1. Introduction to the course. Elementary schemes of data processing. Exploratory data analysis.
2. Simple classification techniques: kNN, decision trees. Classifier quality evaluation.
3. Data classification: Support Vector Machine algorithm, Artificial Neural Networks. Quality of data and its impact on the classification outcome.
4. Ensemble classification: bagging, boosting.
5. Cluster analysis: centroid-based clustering, hierarchical clustering, density-based clustering, Self-Organizing Maps. Evaluation of clustering quality.
6. Regression models: multiple linear regression.
7. Regression models: Generalized Linear Models.
8. Modeling and forecasting of time series.
Laboratories:
Laboratories’ objective is to broaden knowledge of machine learning techniques with a focus on practical abilities. The schedule is parallel to the program of lectures:
1. Exploratory data analysis.
2. Classification.
3. Cluster analysis.
4. Regression models.
5. Time series analysis.
Project classes:
Through the semester students will be carrying on a project work assigned by the teacher. It will be individual project work.
The task will require application of methods discussed during the lectures. It will be based on a given data set. It will be necessary to conduct exploratory data analysis, select appropriate model, tune its parameters, apply it to the data, evaluate and interpret the results. The assignment will be split into a few stages, whose completion on time will be necessary. Each phase will require a progress report covering current stage of advancement. The final stage is delivered together with a final report summarizing project work. In addition, each student presents the results in form of a spoken presentation in front of the class

**Metody oceny:**

Course grade is obtained from the project work and it consists of:
- 40% implementation of the assignment (the solution)
- 30% progress reports
- 30% final report, including evaluation and interpretation of the results
Each phase of the project work has to be completed on time. Delays result in negative points.

**Egzamin:**

nie

**Literatura:**

1. T. Mitchell, Machine Learning, McGraw Hill, 1997.
2. I. H. Witten, E. Frank, M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kauffman, 2011.
3. J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, EXIT, 2005.
4. M. Krzyśko, W. Wołyński, T. Górecki, M. Skorzybut, Systemy uczące się, WNT, 2008.
5. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, 2009.
6. C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
7. R. O. Duda, P. E. Hart, D. G. Stork, Pattern classification, Wiley, 2001.
8. Środowiska: R i RStudio, Python.

**Witryna www przedmiotu:**

e.mini.pw.edu.pl

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W01:**

Knows basic knowledge representation methods for intelligent systems

Weryfikacja:

evaluation of project work

**Powiązane charakterystyki kierunkowe:** DS2\_W03

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W02:**

Knows advanced machine learning methods, evolutionary approaches, and other methods of widely understood computational intelligence

Weryfikacja:

evaluation of project work

**Powiązane charakterystyki kierunkowe:** DS2\_W03

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W03:**

Knows programming languages commonly used in the area of machine learning

Weryfikacja:

evaluation of project work

**Powiązane charakterystyki kierunkowe:**

**Powiązane charakterystyki obszarowe:**