**Nazwa przedmiotu:**

Data Exploration and Visualisation

**Koordynator przedmiotu:**

dr hab. Przemysław Biecek, prof. uczelni

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Data Science

**Grupa przedmiotów:**

Współny

**Kod przedmiotu:**

.

**Semestr nominalny:**

2 / rok ak. 2022/2023

**Liczba punktów ECTS:**

4

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

1. godziny kontaktowe – 65 h; w tym
a) obecność na wykładach – 15 h
b) obecność na zajęciach projektowych – 45 h
c) konsultacje – 5 h
2. praca własna studenta – 45 h; w tym
a) zapoznanie się z literaturą – 10 h
b) przygotowanie projektów – 30 h
c) przygotowanie do zajęć projektowych (prace domowe) – 5 h
Razem 110 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 zajęciach projektowych – 45 h 3. konsultacje – 5 h
Razem 65 h, co odpowiada 3 pkt. ECTS

**Język prowadzenia zajęć:**

angielski

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

1. obecność na zajęciach projektowych – 45 h
2. przygotowanie projektów – 30 h
3. przygotowanie do zajęć projektowych (prace domowe) – 5 h
Razem 75 h, co odpowiada 3 pkt. ECTS

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

|  |  |
| --- | --- |
| Wykład: | 15h |
| Ćwiczenia: | 0h |
| Laboratorium: | 0h |
| Projekt: | 45h |
| Lekcje komputerowe: | 0h |

**Wymagania wstępne:**

Statistics, Programming in R or python

**Limit liczby studentów:**

Number of groups: no limits Project classes – the number of students in a group matches the limits defined by the Warsaw University of Technology

**Cel przedmiotu:**

Learning techniques for data visualization and exploration. Techniques that will be covered: descriptive statistics, statistical tests for group comparison, goodness of fit tests, assessment of the strength of the relationship between variables.
Techniques for visualisation of the distribution relationship between pair and larger group of variables.
Techniques for exploration and explanation of predictive models.

**Treści kształcenia:**

Lecture:
One-dimensional data exploration techniques (location measures, measures of variability)
Two-dimensional data exploration techniques (correlation, statistical tests for the significance of the strength of relations between variables, correspondence analysis, canonical analysis)
Dimension reduction techniques in data exploration (Principal Component Analysis, Multi Dimensional Reduction)
Visualization of single variable distribution (histogram, boxplots) Visualization of variable pair distribution (scatter chart, mosaic chart) Visualization of the distribution of relationships between variable sets (association rules, Bayesian networks)
The topic of testing multiple sets. Problems and challenges in the analysis of high-dimensional data. Techniques of exploration and explainability of the predictive model (model exploration pyramid, bias and fairness topic in bottom analysis)
Project classes:
Development of tools automating data exploration The application of the developed tool for new data from the OpenML

**Metody oceny:**

During the semester, students will have to complete two projects, one related to data visualization and one for data exploration. Projects will be developed at home, but their results will be presented in class. Each of these projects will be evaluated on a scale from 0 to 100 points.
Additional points will be gained from homework and activity in presenting research papers.
To pass it is necessary to obtain a total of at least 50% of points.

**Egzamin:**

nie

**Literatura:**

1. Przemysław Biecek, „Zbiór esejów o sztuce pokazywania danych”, Wydawnictwo Uniwersytetu Warszawskiego, 2016
2. Edward R. Tufte, “The Visual Display of Quantitative Information”, Graphics Press, 2001
3. John Tukey, ,,Exploratory Data Analysis'', Pearson, 1977
We will use R for data exploration and visualisation

**Witryna www przedmiotu:**

.

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W01:**

The graduate knows the key methods, algorithms and Big Data analysis environments as well as various conditions related to data analysis, including ethics

Weryfikacja:

Homeworks + Project 2

**Powiązane charakterystyki kierunkowe:** DS2\_W01

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W02:**

The graduate knows tools for data exploration and communication of results in the area of data analysis

Weryfikacja:

Homeworks + Project 1

**Powiązane charakterystyki kierunkowe:** DS2\_W04

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W03:**

The graduate knows the techniques of clear and correct data presentation using static and interactive graphics and understands the ethical aspects of data presentation

Weryfikacja:

Homeworks + Project 1, 2

**Powiązane charakterystyki kierunkowe:** DS2\_W08

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U01:**

The graduate can prepare statistical plots and interactive visualization of complex data

Weryfikacja:

Homeworks + Project 1

**Powiązane charakterystyki kierunkowe:** DS2\_U06

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U02:**

The graduate knows how to perform exploratory analysis of real data and is able to propose and verify the correctness of the theoretical model

Weryfikacja:

Homeworks + Project 1

**Powiązane charakterystyki kierunkowe:** DS2\_U08

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U03:**

The graduate is able to prepare a holistic solution to a given problem, including data acquisition, initial processing, selection of appropriate methods, e.g. predictive ones and their application, and critical analysis of the obtained results.

Weryfikacja:

Homeworks + Project 2

**Powiązane charakterystyki kierunkowe:** DS2\_U12

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka K01:**

The graduate is able to think and act in a creative and entrepreneurial way.

Weryfikacja:

Homeworks + Project 1, 2

**Powiązane charakterystyki kierunkowe:** DS2\_K04

**Powiązane charakterystyki obszarowe:**