**Nazwa przedmiotu:**

Wstęp do eksploracji danych

**Koordynator przedmiotu:**

dr hab. inż. Przemysław Biecek, prof. PW

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia I stopnia

**Program:**

Matematyka i Analiza Danych

**Grupa przedmiotów:**

Wspólne

**Kod przedmiotu:**

.

**Semestr nominalny:**

4 / rok ak. 2019/2020

**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 – 30 h
b) obecność na laboratoriach – 30 h
d) konsultacje – 5 h
2. praca własna studenta – 50 h; w tym
a) zapoznanie się z literaturą – 8 h
b) rozwiązanie zadań domowych – 20 h
c) przygotowanie do zajęć projektowych – 22 h
Razem 115 h, co odpowiada 5 pkt. ECTS

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

1. obecność na wykładach – 30 h
2. obecność na laboratoriach – 30 h
4. konsultacje – 5 h
Razem 65 h, co odpowiada 2 pkt. ECTS

**Język prowadzenia zajęć:**

polski

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

obecność na laboratoriach – 30 h
rozwiązanie zadań domowych – 20 h
przygotowanie do zajęć projektowych –22 h
Razem 72 h, co odpowiada 3 pkt. ECTS

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

|  |  |
| --- | --- |
| Wykład: | 30h |
| Ćwiczenia: | 0h |
| Laboratorium: | 30h |
| Projekt: | 0h |
| Lekcje komputerowe: | 0h |

**Wymagania wstępne:**

Podstawy programowania i przetwarzania danych
Rachunek prawdopodobieństwa

**Limit liczby studentów:**

.

**Cel przedmiotu:**

Poznanie technik eksploracyjnej analizy danych.
W szczególności technik analizy rozkładu pojedynczych zmiennych oraz par zmiennych.
Wprowadzenie do analizy zależności pomiędzy zmiennymi.
Omówienie technik opisu liczbowego oraz graficznego.
Wykrywanie wartości nietypowych oraz błędów w danych.

**Treści kształcenia:**

Wykład i laboratorium:
1. Cele eksploracyjnej analizy danych
2. Analiza rozkładu jednej zmiennej ciągłej (histogram, wykres pudełko wąsy, jądrowy estymator gęstości)
3. Analiza rozkładu jednej zmiennej dyskretnej (analiza częstości)
4. Analiza zgodności rozkładu (wykres qqplot, rootogram)
5. Analiza wartości nietypowych, skrajnych i błędnych
6. Analiza wartości brakujących
7. Analiza dwóch zmiennych ciągłych (korelacja)
8. Analiza dwóch zmiennych kategorycznych (wykresy mozaikowe)
9. Analiza dwóch zmiennych mieszanych
10. Analiza wielu zmiennych

**Metody oceny:**

Zaliczenie przedmiotu oparte będzie o trzy składowe.
Główna (50%) to prace domowe, których w semestrze będzie 10.
Pozostałe dwie składowe to projekt (30%) oraz aktywność na zajęciach (20%).

**Egzamin:**

nie

**Literatura:**

1. ,,Przewodnik po pakiecie R'', Przemysław Biecek 2017
2. ,,R for Data Science’’ Garrett Grolemund, Hadley Wickham 2018

**Witryna www przedmiotu:**

.

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka WAD\_W01:**

Zna podstawy programowania w języku R. Zna podstawowe metody i algorytmy wykorzystywane w przetwarzaniu danych i przygotowywaniu ich do analizy.

Weryfikacja:

Projekt

**Powiązane charakterystyki kierunkowe:**

**Powiązane charakterystyki obszarowe:**

**Charakterystyka WAD\_W02:**

Zna podstawowe metody eksploracyjnej analizy danych i ich wizualizacji.

Weryfikacja:

Prace domowe

**Powiązane charakterystyki kierunkowe:**

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - umiejętności

**Charakterystyka WAD\_U01:**

Potrafi przygotować dane do analizy, dobrać metodę wizualizacji do specyfiki danych oraz przeprowadzić ich eksploracyjną analizę.

Weryfikacja:

Prace domowe

**Powiązane charakterystyki kierunkowe:**

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka WAD\_K01:**

Potrafi współdziałać i pracować w grupie, przyjmując w niej różne role.

Weryfikacja:

projekt

**Powiązane charakterystyki kierunkowe:** MAD1\_K05

**Powiązane charakterystyki obszarowe:** I.P6S\_KO