

KARTA PRZEDMIOTU OFEROWANEGO W SZKOLE DOKTORSKIEJ

Kod przedmiotu	4606-PS-0000BCJ-0003	Nazwa przedmiotu	w j. polskim	Analiza i wizualizacja danych w języku R		
			w j. angielskim	Data analysis and visualization in R.		
Przynależność do grupy przedmiotów	przedmioty ogólne					
Koordinator przedmiotu	prof. dr hab. inż. Janusz Hołyst, prowadzący: dr inż. Robert Paluch					
Jednostka realizująca	Wyd. Fizyki	Dyscyplina/y naukowa*				
Poziom kształcenia	Kształcenie doktorantów	Semestr	zimowy i letni			
Język zajęć	polski					
Forma zaliczenia:	zaliczenie na ocenę	Sumaryczna liczba godzin w semestrze	30	Sumaryczna liczba ECTS	2	
Minimalna liczba uczestników	6	Maksymalna liczba uczestników	24	Dostępność dla studentów	Nie	
Typ zajęć		Wykład	Ćwiczenia audytoryjne	Ćwiczenia projektowe	Laboratorium	Seminarium
Liczba godzin zajęć	tygodniowo	0	0	0	2	0
	łącznie w semestrze	0	0	0	30	0

* nie dotyczy warsztatu badacza

1. Wymagania wstępne

Brak

2. Cele przedmiotu

Celem przedmiotu jest zapoznanie słuchaczy z podstawami języka R oraz nauczenie ich metod analizy i wizualizacji danych w tym języku.

3. Treści programowe (dla każdego typu zajęć oddzielnie)

Wykład

Nie dotyczy

Laboratorium

1. Wprowadzenie do języka R.

- a) Zmienne, instrukcje sterujące, zapis i odczyt danych do pliku.
- b) Operacje na wektorach i macierzach.
- c) Generacja liczb losowych z zadanego rozkładu.
- d) Połączenie i praca z bazą danych MySQL.

2. Wizualizacja w języku R.

- a) Wykresy XY i skrzynkowe.
- b) Histogramy, binowanie.
- c) Biblioteka ggplot2.
- d) Generacja plików w formatach PNG, JPG i EPS.

3. Wstępne przetwarzanie danych, podstawowe testy i modele statystyczne.

- a) Transformacja danych. Biblioteki tidyr i dplyr.
- b) Testowanie zgodności hipotezy metodami parametrycznymi i nieparametrycznymi.
- c) Dopasowywanie rozkładów teoretycznych do danych eksperymentalnych.

- d) Test chi².
 e) Jedno i wielowymiarowa analiza wariancji.
 f) Regresje liniowa, logistyczna i Poissona.
 4. Eksploracja danych (uczenie maszynowe).
 a) Skuteczność klasyfikacji.
 b) Walidacja krzyżowa.
 c) Analiza dyskryminacyjna.
 d) Analiza skupień.
 e) Skalowanie wielowymiarowe.
 f) Analiza składowych głównych.

4. Efekty uczenia się

Rodzaj efektu	Opis efektu uczenia się	Odniesienie do efektów uczenia się w SD PW	Sposób weryfikacji efektów uczenia*
Wiedza			
W01	Absolwent zna i rozumie podstawy teoretyczne metod statystycznej analizy danych.	SD_W2	Prace domowe /kolokwium pisemne
W02	Absolwent zna główne trendy rozwojowe związane z metodami statystycznej analizy danych.	SD_W3	Prace domowe /kolokwium pisemne
W03			
Umiejętności			
U01	Absolwent potrafi wyciągać poprawne wnioski na podstawie wyników uzyskanych podczas analizy i wizualizacji danych	SD_U1	Prace domowe /kolokwium pisemne
U02			
U03			
Kompetencje społeczne			
K01	Absolwent jest gotów do niezależnego prowadzenia badań naukowych obejmujących analizę i wizualizację danych.	SD_K5	Prace domowe /kolokwium pisemne

* dozwolone sposoby weryfikacji efektów uczenia się: egzamin; egzamin ustny; kolokwium pisemne; kolokwium ustne; ocena projektu; ocena sprawozdania; ocena raportu; ocena prezentacji; ocena aktywności w trakcie zajęć; prace domowe; test

5. Kryteria oceny

Zadania domowe: 10 x 5 pkt = 50 pkt
 Kolokwium końcowe z zadań praktycznych: 20 pkt
 Razem do uzyskania: 70 pkt
 Oceny: 35-41 dst, 42-48 dst+, 49-55 db, 56-62 db+, 63-70 bdb

6. Literatura

Literatura podstawowa:

- [1] P. Biecek, Przewodnik po pakiecie R, Oficyna Wydawnicza Gis, Wrocław 2017.
 [2] T. Górecki, Podstawy statystyki z przykładami w R, Wydawnictwo BTC, Legionowo 2011.
 [3] M. Walesiak, E. Gatnar, Statystyczna analiza danych z wykorzystaniem pakietu R, Wydawnictwo Naukowe PWN, Warszawa 2013.

Literatura uzupełniająca:

- [1] K. Seefeld, E. Linder, Statistics Using R with Biological Examples, https://cran.r-project.org/doc/contrib/Seefeld_StatsRBio.pdf
 [2] A. Coghlan, A Little Book of R For Biomedical Statistics, <http://a-little-book-of-r-for-biomedical-statistics.readthedocs.io/en/latest/>
 [3] A. Coghlan, A Little Book of R for Multivariate Analysis, <http://little-book-of-r-for-multivariate-analysis.readthedocs.io/en/latest/>

7. Nakład pracy doktoranta niezbędny do osiągnięcia efektów uczenia się**		
Lp.	Opis	Liczba godzin
1	godziny kontaktowe z nauczycielem akademickim wynikające z planu	30
2	Godziny kontaktowe z nauczycielem akademickim w ramach konsultacji, egzaminów, sprawdzianów itp.	5
3	Godziny pracy samodzielnej doktoranta w ramach przygotowania do zajęć oraz opracowania sprawozdań, projektów, prezentacji, raportów, prac domowych	15
4	godziny pracy samodzielnej doktoranta w ramach przygotowania do egzaminu, sprawdzianu, zaliczenia	5
Sumaryczny nakład pracy doktoranta		55
Liczba punktów ECTS		2

** 1 ECTS pracy = 25-30 godzin nakładu pracy doktoranta (np. 2 ECTS = 60 godzin; 4 ECTS = 110 godzin)