

SYLABUS PRZEDMIOTU W SZKOLE DOKTORSKIEJ

Tytuł	<i>Analiza danych przy wykorzystaniu języka R</i>
Tytuł w jęz. ang.	Data analysis using R

Status przedmiotu	obowiązkowy dla: <i>wspierający SzD nauki o zarządzaniu i jakości</i>
	do wyboru dla:

Autor/autorzy sylabusa: Mariusz Rafało	Zespół :	koordynator:
		członek zespołu
		członek zespołu

Sygnatura przedmiotu:

Część A

1. Syntetyczna charakterystyka przedmiotu (*główne hasła – około 400 znaków*): Zagadnienia poruszane w trakcie zajęć koncentrują się na pracy z danymi. W trakcie zajęć studenci poznają techniki i narzędzia służące do analizy i wizualizacji danych. Przedstawione zostaną zasady analizy eksploracyjnej (rozkład zmiennych, korelacja, statystyki opisowe) oraz zasady analizy mającej na celu wyjaśnienie zjawisk (predykcja zjawisk, analiza anomalii, podział zbiorowości na segmenty, itp.). Studenci zostaną zapoznani z metodami oceny i porównań różnych modeli, realizujących ten sam cel.

2. Słowa kluczowe (*3 – 6 słów*): r, analiza danych, modele predykcyjne, ROC, regresja, k-means

Część B

Przedmiotowe efekty uczenia się		Powiązanie z efektami uczenia się dla SzD
Wiedza (<i>liczba efektów od 2 do 5</i>)		
	Po zrealizowaniu programu przedmiotu student powinien:	
W.1	Znać składnię języka R	
W.2	Znać podstawowe narzędzia języka R, które pozwalają na przeprowadzenie analizy eksploracyjnej	

W.3	Znać metody i narzędzia analizy i wizualizacji danych w języku R	
W.4	Znać techniki budowy modeli: drzewa decyzyjne, regresja logistyczna, podział na klastry, asocjacje	
W.5	Znać metody oceny jakości modeli: ROC, macierz pomyłek	
Umiejętności (liczba efektów od 2 do 5)		
U.	Po zrealizowaniu programu przedmiotu student potrafi:	
U.1	Napisać program w języku R, przeprowadzający analizę i wizualizację danego zbioru danych	
U.2	Dokonać połączenia kilku zbiorów danych, agregowania i filtrowania tych danych, względem zadanych warunków	
U.3	Dokonać analizy wyjaśniającej w oparciu o drzewo decyzyjne lub regresję logistyczną	
U.4	Dokonać podziału zbioru na klastry za pomocą metody k-średnich	
U.5	Dokonać obiektywnej ewaluacji zbudowanego modelu	
Kompetencje społeczne (liczba efektów od 1 do 3)		
	Po zrealizowaniu programu przedmiotu student wykazuje:	
K.1	Umiejętność komunikowania zagadnień związanych z analizą danych	
K.2	Umiejętność komunikowania zagadnień związanych z programowaniem w języku R	

Część C

Semestralny plan zajęć:

1. Wprowadzenie do języka R (Składnia, typy danych, zmienne, itp.)
2. Struktury danych (dataframe, matrix, factor, list)

3. Typy danych, funkcje, instrukcje warunkowe i pętle

4. Praca z danymi

- * faktoryzacja
- * podział na przedziały
- * braki danych
- * transformacje danych
- * podstawowe statystyki opisowe
- * weryfikacja normalności rozkładu

5. Operacje na danych: agregacja, grupowanie, filtrowanie

6. Podstawowa analiza i wizualizacja danych

7. Budowa modeli ekonometrycznych

8. Modele nadzorowane i nienadzorowane

9. Ocena jakości modeli (ROC)

Literatura podstawowa (jeśli wybrane fragmenty publikacji zwartych, to wskazane podanie rozdziałów, ew. stron):

1. J. Maindonald, Data analysis and graphics using R: an example-based approach, Cambridge UK, New York: Cambridge University Press, 2003
2. G. Golemund and H. Wickham, R for Data Science, O'Reilly, 2018

Literatura uzupełniająca (jeśli wybrane fragmenty publikacji zwartych, to wskazane podanie rozdziałów, ew. stron):

1. G. James, D. Witten, T. Hastie, and R. Tibshirani, Introduction to Statistical Learning, Springer, 2013

Część D	
Forma zajęć:	Wymiar zajęć w godz.:
Ogółem godzin <i>w tym:</i>	30
Wykład	10
Ćwiczenia	20
Elementy oceny końcowej (ogółem 100%), w tym:	
projekt	100%

Liczba punktów ECTS	3
----------------------------	---

Część E
Metody dydaktyczne (nauczania) stosowane przez prowadzącego
M.2. wykład z wykorzystaniem technik multimedialnych M.5. laboratorium komputerowe

Część F
Metody weryfikacji (sprawdziany) osiągnięcia przedmiotowych efektów kształcenia
W.7. projekt