



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Wprowadzenie do programowania w R [S1DSwB1>WdPwR]

Przedmiot

Kierunek studiów

Data Science w biznesie

Rok/Semestr

3/5

Studia w zakresie (specjalność)

–

Profil studiów

ogólnoakademicki

Poziom studiów

pierwszego stopnia

Język oferowanego przedmiotu

polski

Forma studiów

stacjonarne

Wymagalność

obligatoryjny

Liczba godzin

Wykład

30

Laboratorium

0

Inne

0

Ćwiczenia

30

Projekty/seminaria

0

Liczba punktów ECTS

5,00

Koordynatorzy

dr hab. inż. Yevhen Revtiuk

yevhen.revtiuk@put.poznan.pl

Wykładowcy

Wymagania wstępne

Podstawowa znajomość obsługi komputera. Umiejętność instalowania i uruchamiania oprogramowania. Podstawowe operacje na plikach i katalogach (np. zapisywanie, przenoszenie, kopiowanie). Znajomość podstawowych działań matematycznych. Rozumienie pojęć takich jak zmienne, funkcje, tablice danych. Logiczne myślenie i umiejętność rozwiązywania problemów. Umiejętność analizowania prostych problemów i ich strukturyzowania.

Cel przedmiotu

Celem przedmiotu jest wprowadzenie studentów do podstaw programowania w języku R, ze szczególnym uwzględnieniem analizy i wizualizacji danych. Studenci nauczą się podstaw składni języka, pracy z danymi oraz wykorzystywania pakietów R do rozwiązywania problemów obliczeniowych i statystycznych. Kurs przygotowuje uczestników do samodzielnej pracy z R w kontekście inżynierii, nauk ścisłych i analizy danych.

Przedmiotowe efekty uczenia się

Wiedza:

1. Charakteryzuje podstawowe metody analizy danych w języku R, w tym operacje na strukturach danych, programowanie oraz wizualizację wyników [DSB1_W01].

2. Opisuje zagadnienia związane z przetwarzaniem danych w języku R, w tym importowanie, czyszczenie i transformację zbiorów danych [DSB1_W02].
3. Wyjaśnia metody statystyczne i techniki eksploracyjnej analizy danych dostępne w pakiecie R oraz ich zastosowanie w analizie biznesowej [DSB1_W03].
4. Przedstawia techniki wizualizacji danych oraz narzędzia do tworzenia wykresów za pomocą ggplot2 i innych bibliotek [DSB1_W07].
5. Charakteryzuje podstawowe modele ekonometryczne i statystyczne implementowane w środowisku R, w tym regresję liniową i analizę korelacji [DSB1_W09].

Umiejętności:

1. Projektuje i implementuje analizy danych w R, wykorzystując operacje na wektorach, ramkach danych i listach [DSB1_U03].
2. Analizuje i wizualizuje dane pochodzące z różnych źródeł, stosując funkcje statystyczne i graficzne dostępne w R [DSB1_U04].
3. Formułuje specyfikacje problemów analitycznych oraz dobiera odpowiednie metody eksploracji i analizy danych w R [DSB1_U05].
4. Argumentuje wybór narzędzi i metod analitycznych w środowisku R oraz ocenia skuteczność zastosowanych rozwiązań [DSB1_U11].
5. Rozwija swoje kompetencje w zakresie programowania w R, analizując dokumentację i korzystając z dostępnych zasobów edukacyjnych [DSB1_U15].

Kompetencje społeczne:

1. Krytycznie analizuje własne umiejętności programistyczne i statystyczne, dążąc do ich rozwoju i aktualizacji w kontekście nowych metod analizy danych [DSB1_K01].
2. Korzysta z aktualnego dorobku naukowego w dziedzinie analizy danych i programowania w R, uwzględniając jego praktyczne zastosowanie w biznesie [DSB1_K02].

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

1. Ocena projektów końcowych (50%)

• Opis: Studenci przygotowują projekt końcowy, w którym tworzą kompleksową analizę danych, wykorzystując język R. Projekt musi obejmować następujące elementy: pozyskiwanie danych, ich wstępną obróbkę (czyszczenie danych), analizę statystyczną oraz przedstawienie wyników w formie wykresów i raportu. Zaletą projektu będzie również umiejętność zastosowania odpowiednich bibliotek i funkcji w R w zależności od problemu analitycznego.

• Kryteria oceny:

- Jakość i poprawność analizy danych: Poprawność metod analitycznych, skuteczność czyszczenia i przygotowania danych, poprawne dobieranie technik analitycznych w zależności od rodzaju danych.
- Umiejętność programowania w R: Znajomość składni, użycie funkcji i pakietów R w kontekście rozwiązania zadania, struktura kodu, czytelność i efektywność kodu.
- Oryginalność podejścia i kreatywność: Innowacyjne podejście do analizy danych, kreatywne zastosowanie narzędzi programistycznych.
- Przejrzystość raportu i komunikacja wyników: Jasne przedstawienie wyników, przejrzystość analizy, uzasadnienie zastosowanych metod i wykresów, poprawność i spójność raportu.

2. Ćwiczenia praktyczne i zadania domowe (50%)

• Opis: Regularne zadania domowe oraz ćwiczenia praktyczne podczas zajęć, polegające na rozwiązywaniu problemów programistycznych i analizie danych w R. Zadania będą obejmować zarówno podstawowe operacje na danych (wczytywanie, czyszczenie), jak i bardziej zaawansowane zagadnienia, takie jak analiza statystyczna czy wizualizacja danych.

• Kryteria oceny:

- Dokładność wykonania zadań: Poprawność rozwiązywanych zadań, odpowiednia implementacja algorytmów, poprawność wyników obliczeń.
- Zastosowanie odpowiednich funkcji w R: Umiejętność wyboru i zastosowania odpowiednich narzędzi i funkcji w zależności od zadania (np. pakiety dplyr, ggplot2, tidyr).
- Terminowość i samodzielność: Terminowe oddanie zadań oraz samodzielność w rozwiązywaniu problemów programistycznych.
- Umiejętność przedstawienia wyników: Umiejętność przedstawienia wyników analizy w formie wykresów, tabel lub raportów, dostosowanych do odbiorcy (np. wykresy na potrzeby analizy vs. wykresy do prezentacji).

Treści programowe

1. Wprowadzenie do języka R
 - Krótka historia i zastosowania języka R.
 - Instalacja i konfiguracja R oraz RStudio.
 - Podstawowe operacje w konsoli R.
2. Podstawowe elementy składni języka R
 - Typy danych (liczbowe, logiczne, znakowe).
 - Operatory arytmetyczne, logiczne i przypisania.
 - Zmienne i podstawowe operacje na nich.
3. Struktury danych w R
 - Wektory i operacje na wektorach.
 - Macierze i tablice.
 - Ramki danych (data frames) - podstawowe operacje, filtrowanie, sortowanie.
 - Listy - struktura i zastosowanie.
4. Praca z danymi w R
 - Wczytywanie i eksportowanie danych (CSV, Excel, bazy danych).
 - Podstawowe operacje na zbiorach danych (łączenie, filtrowanie, agregacja).
 - Obsługa brakujących wartości i czyszczenie danych.
5. Programowanie w R
 - Instrukcje warunkowe (if, else).
 - Pętle (for, while) i alternatywne rozwiązania (apply, lapply, sapply).
 - Definiowanie własnych funkcji.
6. Wizualizacja danych
 - Podstawowe wykresy w R (histogramy, wykresy pudełkowe, punktowe).
 - Tworzenie wykresów za pomocą pakietu ggplot2.
 - Personalizacja wykresów i eksport wyników.
7. Podstawy analizy danych i statystyki w R
 - Podstawowe statystyki opisowe (średnia, mediana, odchylenie standardowe).
 - Wykorzystanie funkcji statystycznych i testów istotności.
 - Analiza korelacji i regresji liniowej.
8. Mini-projekt końcowy
 - Samodzielne rozwiązanie problemu analizy danych w R.
 - Przedstawienie wyników w formie raportu lub prezentacji.

Tematyka zajęć

1. Wprowadzenie do R i środowiska RStudio
 - Instalacja i konfiguracja R oraz RStudio.
 - Pierwsze kroki w R - podstawowe operacje w konsoli.
 - Podstawowa składnia języka R.
2. Typy i struktury danych w R
 - Podstawowe typy danych (liczbowe, logiczne, znakowe).
 - Wektory, macierze, tablice - tworzenie i operacje.
 - Ramki danych (data frames) - filtrowanie, sortowanie, modyfikacje.
 - Listy i ich zastosowanie.
3. Podstawowe operacje na danych
 - Importowanie i eksportowanie danych (CSV, Excel, bazy danych).
 - Czyszczenie i przekształcanie danych.
4. Programowanie w R
 - Instrukcje warunkowe (if, else).
 - Pętle (for, while) i alternatywne metody iteracji (apply, lapply, sapply).
 - Tworzenie własnych funkcji.
5. Wizualizacja danych w R
 - Tworzenie podstawowych wykresów (histogramy, wykresy punktowe, pudełkowe).
 - Pakiet ggplot2 - zaawansowana wizualizacja danych.
 - Personalizacja wykresów i eksport wyników.
6. Podstawowe statystyki i analiza danych
 - Statystyki opisowe (średnia, mediana, wariancja, odchylenie standardowe).
 - Analiza korelacji i testy statystyczne.
 - Regresja liniowa w R.

7. Zaawansowana analiza danych (opcjonalnie, dla bardziej zaawansowanych grup)
 - Klasteryzacja i analiza skupień.
 - Wprowadzenie do regresji wielorakiej i modeli predykcyjnych.
 - Przetwarzanie danych na dużą skalę.
8. Równocześnie wplatanie niezbędnych wiadomości ze statystyki stanowiących podstawę analiz w pakiecie r.

Metody dydaktyczne

1. Wykłady teoretyczne

- Podczas wykładów omawiane będą kluczowe zagadnienia związane z programowaniem w R, w tym składnia języka, podstawowe typy danych, struktury danych, manipulacja danymi oraz podstawy analizy statystycznej. Wykłady będą wprowadzać studentów w teoretyczne aspekty pracy w R oraz przedstawiać zasady, które będą wykorzystywane w ćwiczeniach praktycznych.

2. Ćwiczenia praktyczne

- Zajęcia praktyczne będą głównie skupiały się na pracy z językiem R, gdzie studenci samodzielnie wykonają ćwiczenia związane z analizą danych, implementacją algorytmów i wizualizacją wyników. Będą to zadania indywidualne oraz w grupach, które pozwolą na praktyczne zastosowanie poznanych teorii.
- Studenci będą mieli okazję na bieżąco stosować omawiane zagadnienia, rozwiązując konkretne problemy związane z przetwarzaniem danych w R.

3. Laboratoria komputerowe

- Zajęcia komputerowe, w których studenci będą pracować na swoich laptopach, pozwolą im na bezpośrednią pracę z językiem R. Każdy student będzie miał dostęp do niezbędnego oprogramowania, aby móc samodzielnie rozwiązywać problemy i zadania. Instruktorzy będą dostępni do pomocy i konsultacji.
- Celem laboratoriów jest umożliwienie studentom praktycznego opanowania umiejętności programowania w R oraz nauka efektywnej pracy z danymi.

4. Konsultacje indywidualne

- Studenci będą mogli skorzystać z konsultacji indywidualnych z prowadzącym zajęcia w celu omówienia trudności związanych z materiałem lub zadaniami. Konsultacje będą dotyczyły zarówno teorii, jak i praktycznych aspektów programowania w R.

5. Studium przypadków i analiza danych rzeczywistych

- Na zajęciach zostaną omówione przykłady rzeczywistych danych, które studenci będą analizować w kontekście różnych zagadnień programistycznych. Studium przypadków pozwoli na zastosowanie nabytych umiejętności do rozwiązywania realnych problemów i zwiększy motywację studentów do nauki.
- Prezentacje i dyskusje
- Podczas zajęć studenci będą prezentować wyniki swoich projektów lub rozwiązań zadań, co umożliwi im doskonalenie umiejętności komunikacji wyników oraz obrony swoich rozwiązań. Będą również brać udział w dyskusjach na temat wybranych zagadnień związanych z programowaniem i analizą danych w R.

Literatura

Podstawowa:

1. Crawley, M. J. (2013). The R Book. Wiley.
2. Kabacoff, R. I. (2015). R in Action: Data Analysis and Graphics with R. Manning.
3. Wickham, H., & Grolemund, G. (2017). R for Data Science. O'Reilly Media.
4. Zuur, A. F., Ieno, E. N., & Meesters, E. H. W. G. (2009). A Beginner's Guide to R. Springer.

Uzupełniająca:

1. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer.
2. Matloff, N. (2011). The Art of R Programming: A Tour of Statistical Software Design. No Starch Press.
3. Xie, Y. (2016). R Markdown: The Definitive Guide. CRC Press.
4. Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S. Springer.
5. Peng, R. D. (2016). R Programming for Data Science. Leanpub.

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	125	5,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	62	2,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwiiw/egzaminu, wykonanie projektu)	63	2,50