

## KARTA KURSU

Nazwa	<b>Inżynieria i analiza danych</b>
Nazwa w j. ang.	Data science

Koordynator	dr Roman Czapla	Zespół dydaktyczny
		dr Roman Czapla mgr inż. Katarzyna Marczak dr Wojciech Nawalaniec
Punktacja ECTS*	st. stacjonarne: 3 st. niestacjonarne: 2	

### Opis kursu (cele kształcenia)

Celem kursu jest zapoznanie studenta ze współczesnymi metodami analizy danych.

### Warunki wstępne

Wiedza	Podstawowa wiedza o programowaniu strukturalnym.
Umiejętności	Umiejętność pisania prostych skryptów (np. w języku Python).
Kursy	Podstawy programowania, Programowanie proceduralne, Wstęp do matematyki, Matematyka 1, Matematyka 2, Metody numeryczne, Języki skryptowe

### Efekty kształcenia

	Efekt kształcenia dla kursu	Odniesienie do efektów kierunkowych
Wiedza	Po zakończeniu kursu student: W01: zna podstawy wnioskowania statystycznego i programowania matematycznego; W02: rozumie zasadę działania i stosowania zagadnień oraz algorytmów związanych ze współczesną analizą danych; W03: zna popularne moduły języka Python wykorzystywane w analizie danych.	K_W06  K_W05  K_W13

Umiejętności	Efekt kształcenia dla kursu	Odniesienie do efektów kierunkowych
	Po zakończeniu kursu student: U01: potrafi przeprowadzić weryfikację hipotezy statystycznej; U02: potrafi przeprowadzić analizę danych celem wydobycia z nich wiedzy umożliwiającej zrozumienie opisywanego przez te dane zjawiska lub poprawę jakości procesu decyzyjnego; U03: potrafi wykorzystać popularne moduły języka Python wykorzystywane w analizie danych; U04: Potrafi wykorzystać podstawowe narzędzia programowania matematycznego.	K_U06 K_U06, K_U02, K_U13 K_U04 K_U06
Kompetencje społeczne	Efekt kształcenia dla kursu	Odniesienie do efektów kierunkowych
	Po zakończeniu kursu student: K01: jest świadomy korzyści płynących z umiejętności rozwiązywania problemów z zastosowaniem analizy danych.	K_K01

### Studia stacjonarne

		Organizacja										
Forma zajęć	Wykład (W)	Ćwiczenia w grupach										
		A		K		L		S		P		E
Liczba godzin						30						

### Studia niestacjonarne

		Organizacja										
Forma zajęć	Wykład (W)	Ćwiczenia w grupach										
		A		K		L		S		P		E
Liczba godzin						20						

### Opis metod prowadzenia zajęć

Prowadzący przedstawia nowe zagadnienia oraz problemy wraz z ewentualnymi sugestiami dot. Możliwych metod ich rozwiązania. Studenci piszą skrypt rozwiązujący zadany problem. Następnie odbywa się wspólna analiza rozwiązania.

## Formy sprawdzania efektów kształcenia

	E – learning	Gry dydaktyczne	Ćwiczenia w szkole	Zajęcia terenowe	Praca laboratoryjna	Projekt indywidualny	Projekt grupowy	Udział w dyskusji	Referat	Praca pisemna (esej)	Egzamin ustny	Egzamin pisemny	Inne
W01					X			X					
W02					X			X					
W03					X			X					
U01					X			X					
U02					X			X					
U03					X			X					
U04					X			X					
K01								X					

Kryteria oceny	<p>Podstawą do zaliczenia jest ocena uzyskana na podstawie oddanych zadań związanych z analizą danych (alternatywnie: na podstawie testu przygotowanego przez prowadzącego).</p> <p>Ocenę dobrą lub bardzo dobrą może uzyskać student, który otrzymał odpowiednią liczbę punktów z określonych (alternatywnie: uzyskał odpowiednią liczbę punktów na drodze testu przygotowanego przez prowadzącego).</p>
----------------	---

Uwagi	
-------	--

## Treści merytoryczne (wykaz tematów)

<ol style="list-style-type: none"> <li>1. Biblioteki NumPy i Pandas <ul style="list-style-type: none"> <li>• wektory, macierze i inne tablice NumPy <ul style="list-style-type: none"> <li>– tworzenie i reprezentacja tablic</li> <li>– podstawowe metody i funkcje</li> <li>– indeksowanie tablic</li> </ul> </li> <li>• struktury danych biblioteki Pandas <ul style="list-style-type: none"> <li>– ramki danych</li> <li>– obiekty typu Series</li> <li>– obiektu typu index</li> <li>– indeksowanie</li> <li>– wybrane operacje</li> </ul> </li> </ul> </li> <li>2. Elementy statystyki z wykorzystaniem języka Python <ul style="list-style-type: none"> <li>• atrybuty i ich typy</li> <li>• pomiar tendencji centralnej</li> <li>• pomiar dyspersji</li> <li>• skośność i kurtoza</li> <li>• określanie związków za pomocą współczynników kowariancji i korelacji</li> <li>• centralne twierdzenie graniczne 96</li> <li>• pozyskiwanie prób</li> <li>• wybrane rozkłady prawdopodobieństwa</li> <li>• estymacje parametrów i charakterystyk rozkładów</li> <li>• testy statystyczne</li> </ul> </li> <li>3. Elementy algebry liniowej <ul style="list-style-type: none"> <li>• macierze i wyznaczniki</li> <li>• równania liniowe</li> <li>• wartości i wektory własne</li> </ul> </li> </ol>	
--	--

- generowanie liczb losowych
- 4. Odczytywanie i zapisywanie danych
  - odczyt i zapis danych w formacie tekstowym
  - formaty danych binarnych
  - obsługa interfejsów sieciowych
  - obsługa baz danych
- 5. Czyszczenie i przygotowywanie danych
  - obsługa brakujących danych
  - przekształcanie danych
  - rozszerzone typy danych
  - techniki kodowania, skalowania, przekształcania i rozdzielania cech
- 6. Wizualizacja danych
  - podstawy obsługi pakietu matplotlib
  - wykresy z pakietami pandas i seaborn
  - inne narzędzia przeznaczone do wizualizacji danych w języku Python
- 7. Agregacja danych i operacje na grupach
- 8. Szeregi czasowe
- 9. Uczenie nadzorowane: analiza regresyjna
  - regresja liniowa
  - regresja wielomianowa
  - regresja logistyczna
- 10. Uczenie nadzorowane: techniki klasyfikacji
  - klasyfikacja
  - naiwny klasyfikator Bayesa
  - drzew decyzyjne
  - maszyny wektorów nośnych
  - podział danych – zestaw uczący i testowy
  - ocena skuteczności modelu klasyfikacji
- 11. Uczenie nienadzorowane: PCA i analiza skupień
  - redukcja wymiarowości danych
  - analiza głównych składowych
  - analiza skupień
  - widmowa analiza skupień
  - algorytm DBSCAN
  - ocena jakości analizy skupie

#### Wykaz literatury podstawowej

1. W. McKinney, *Python w analizie danych. Przetwarzanie danych za pomocą pakietów pandas i NumPy oraz środowiska Jupyter. Wydanie III*, Helion 2023 (wybrane fragmenty);
2. A.h Navlani, A. Fandango, I. Idris, *Python i praca z danymi. Przetwarzanie, analiza, modelowanie i wizualizacja. Wydanie III*, Helion 2022 (wybrane fragmenty);
3. T. Nield, *Podstawy matematyki w data science. Algebra liniowa, rachunek prawdopodobieństwa i statystyka*, Helion 2023 (wybrane fragmenty);
4. R. Johansson, *Matematyczny Python. Obliczenia naukowe i analiza danych z użyciem NumPy, SciPy i Matplotlib*, Helion 2021; (wybrane fragmenty);

#### Wykaz literatury uzupełniającej

1. M. Walker, *Czyszczenie danych w Pythonie. Receptury. Nowoczesne techniki i narzędzia Pythona do wykrywania i eliminacji zanieczyszczeń oraz wydobywania kluczowych cech z danych*, Helion 2021;
2. A. Boschetti, L. Massaron, *Python. Podstawy nauki o danych. Wydanie II*, Helion 2017;
3. M. Gągolewski, N. Bartoszek, A. Cena, *Przetwarzanie i analiza danych w języku Python*, PWN 2016 (wybrane fragmenty).

Bilans godzinowy zgodny z CNPS (Całkowity Nakład Pracy Studenta) **studia stacjonarne**

liczba godzin w kontakcie z prowadzącymi	Wykład	
	Konwersatorium (ćwiczenia, laboratorium itd.)	30
	Pozostałe godziny kontaktu studenta z prowadzącym	5
liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	20
	Przygotowanie krótkiej pracy pisemnej lub referatu po zapoznaniu się z niezbędną literaturą przedmiotu	
	Przygotowanie projektu lub prezentacji na podany temat (praca w grupie)	
	Przygotowanie do egzaminu/zaliczenia	20
Ogółem bilans czasu pracy		75
Liczba punktów ECTS w zależności od przyjętego przelicznika		3

Bilans godzinowy zgodny z CNPS (Całkowity Nakład Pracy Studenta) **studia niestacjonarne**

liczba godzin w kontakcie z prowadzącymi	Wykład	
	Konwersatorium (ćwiczenia, laboratorium itd.)	20
	Pozostałe godziny kontaktu studenta z prowadzącym	2
liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	15
	Przygotowanie krótkiej pracy pisemnej lub referatu po zapoznaniu się z niezbędną literaturą przedmiotu	
	Przygotowanie projektu lub prezentacji na podany temat (praca w grupie)	
	Przygotowanie do egzaminu/zaliczenia	13
Ogółem bilans czasu pracy		50
Liczba punktów ECTS w zależności od przyjętego przelicznika		2